CAHOOTS Arrival and Clear times

by Luke Warriner

Background

Crisis Assistance Helping Out On The Streets (CAHOOTS) is a service that can be deployed alongside or instead of the police to respond to 911 calls. They are equipped to handle medical and mental health emergencies. They have a maximum of two vans in service at any one time. There have been many complaints about CAHOOTS response times, claiming that they are very slow in responding and arriving to calls. This leads to the primary research question of this project; how EPD and CAHOOTS arrival times and clear times vary by call priority, by time of day, by day of week, and by call type.

Data

The data for this project was gathered from a public records request and is formatted in 12 csv files. Each csv file contains computer aided dispatch (CAD) call data from 2014 to early 2025. There are 20 columns in each csv file. The following steps were taken to clean the data:

- Combine the 12 csv files into 1 csv file
- Change the format of dispatch, arrival, and clear times from seconds to minutes
- Create a new column that identifies whether the call was handled by CAHOOTS or EPD
- Extract the hour and day of the week from the time the call was placed
- Drop any rows where dispatch, arrival and clear time having missing or negative values

The strength of this dataset is that it contains a large number of calls (around 1.4 million), so the results will be representative. A caveat of this data is that it only displays the primary unit for each call. This means that if CAHOOTS and EPD both responded to a call, but EPD was set as the primary unit, information on CAHOOTS for that call may be missing.

*A more in-depth list of steps and a link to GitHub containing scripts and files will be at the end of this report

Methods

The primary test to see if there is a difference between EPD and CAHOOTS will be an independent t-test. The t-test serves to determine if any differences between EPD and CAHOOTS can be explained by random chance or if the difference is statistically significant. the programing language used in this project will be Python and the following packages for Python will be used as well:

- Pandas
- Seaborn
- Scipy.stats
- Matplotlib.pyplot

Each of the following steps was performed using both arrival times and clear times (labeled "time type" in steps below to avoid repetition).

- Seaborn line plot of time type by hour of day separated by service (EPD and CAHOOTS) with call source "SELF" included (All others have it excluded as explained later)
- Seaborn line plot of time type by hour of day separated by service (EPD and CAHOOTS)
- Seaborn line plot of time type by day of week separated by service
- Seaborn bar plot of time type by call type separated by service
- Seaborn bar plot of time type by top 15 call natures (for CAHOOTS) separated by service

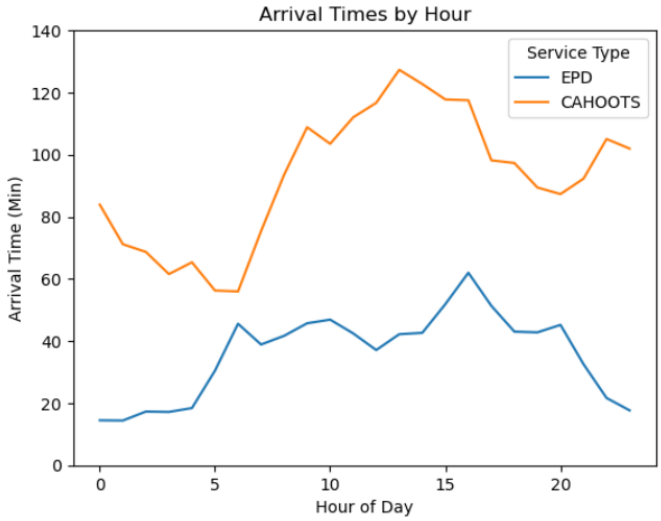
Additionally, for each hour, day, and call type, a t-test was performed to see if the difference between CAHOOTS and EPD was statistically significant. Since the majority of findings are statistically significant, unless stated or labeled otherwise, assume statistical significance in results.

The following was performed on just the subset of calls that CAHOOTS responded to

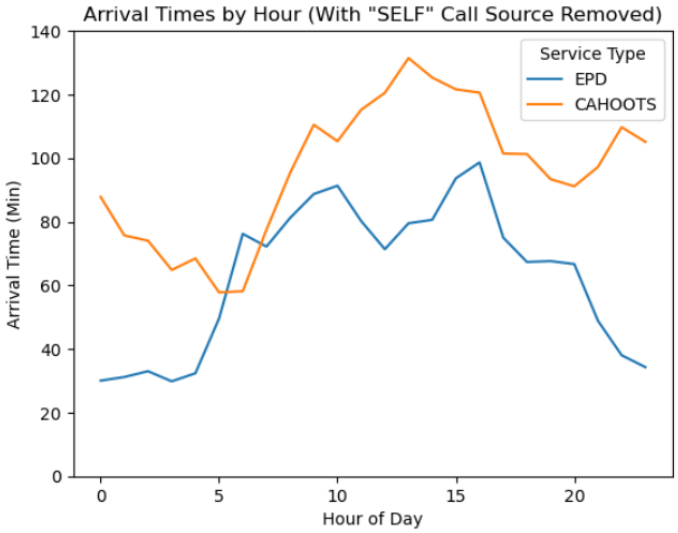
- Seaborn regression plot of arrival times by call volume of each hour in the data set
- Seaborn regression plot of log arrival times by call volume of each hour in the data set

*As with the data section, a more in-depth list of steps and a link to GitHub containing scripts and files will be at the end of this report

Results



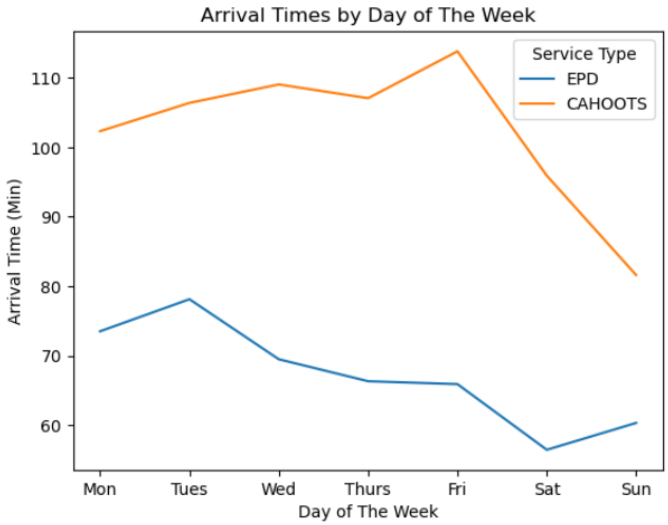
This is a graph of arrival times by hour of the day. The arrival times of EPD are in blue and CAHOOTS is in orange. These colors will remain consistent any time EPD and CAHOOTS are being compared in a graph. CAHOOTS is always slower to arrive to calls than EPD, but the difference is large from the times 9am-3pm. Since the difference is so large and consistent across hours, exploring some possible reasons for this has some merit.



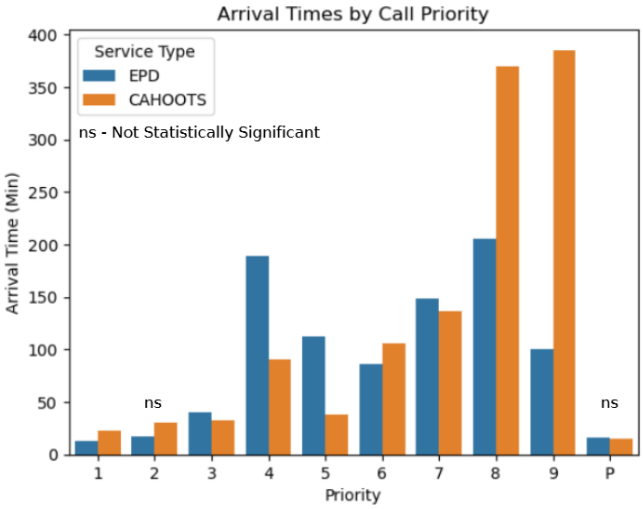
The call source “SELF” is used whenever the unit responding to a call is the one to make the call. This usually means that the unit has seen something and is already on or near the scene. For example, an EPD officer seeing someone speeding and then making a traffic stop. The average arrival time for “SELF” calls is 1.1 minutes and 98.4% of these calls were responded to by EPD. If a unit is already essentially on the scene, calculating arrival time is

not relevant. Additionally, CAHOOTS units do not typically function to intervene on their own accord. They typically respond to calls from CAD and do not patrol, so comparing them to EPD under the same conditions is more appropriate. When the "SELF" call source is removed, the CAHOOTS arrival times do not noticeably change, but EPD's response time increases noticeably. CAHOOTS is still slower over the majority of hours (except for 6am), but the difference is much smaller with the "SELF" call source removed.

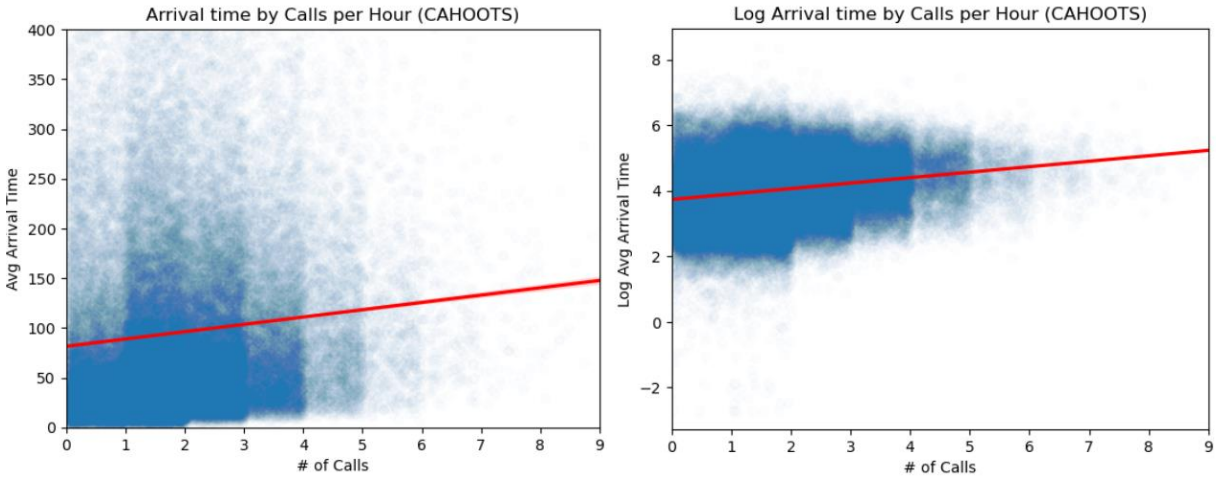
*For all future graphs, the call source "SELF" will be removed for the reasons mentioned above



When comparing arrival times over days of the week, CAHOOTS is always slower than EPD by roughly 30-40 minutes. EPD arrival times peak on Tuesday and then fall until Sunday when they rise again. CAHOOTS on the other hand rise until Friday then fall until Monday, when they rise again.

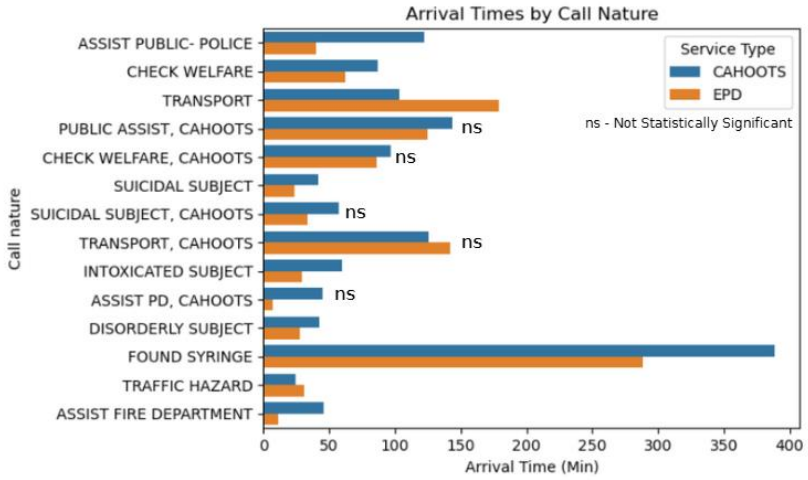


When looking at arrival times by call priority, priority 2 and priority P are not statistically significant, which means there is not enough evidence to say that there is a difference between EPD and CAHOOTS for those two priorities. EPD is faster for priorities 1,6,8, and 9, while CAHOOTS is faster for priorities 4, 5, and 7. It should be noted that the majority of calls that CAHOOTS responds to are priority 7. It is possible that CAHOOTS responds slower to priorities that are unusual for them call volume wise, but this would require further investigation. Despite the low number of calls, the difference in arrival time between EPD and CAHOOTS is very large for priorities 8 and 9. It is possible that if CAHOOTS is at full capacity, when 2 calls come in and they must decide which call to respond to, they will likely respond to the more important call. This means that the less important calls (such as priorities 8 and 9) could get pushed back multiple times if higher priority calls come in.

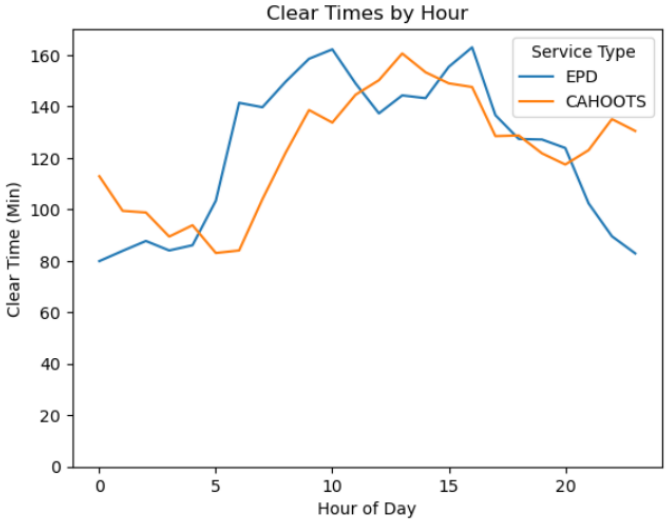


The number of calls in each hour is based on the time the call was placed. As this number of calls increases, so does CAHOOTS response time. The graph on the left shows that the difference between 1 call in an hour and 4 calls has an increase of around 30 minutes to response times.

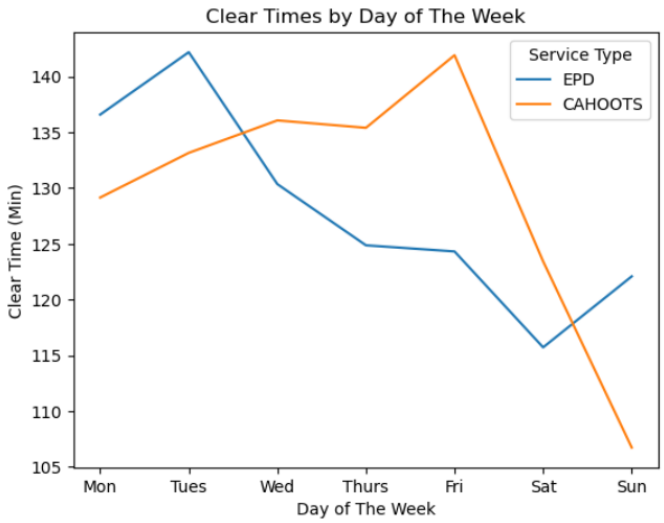
*The left graph does not display some of the higher values in order to display the change in regression line better, so the right graph has been added with log time on the y axis to show the full distribution of calls



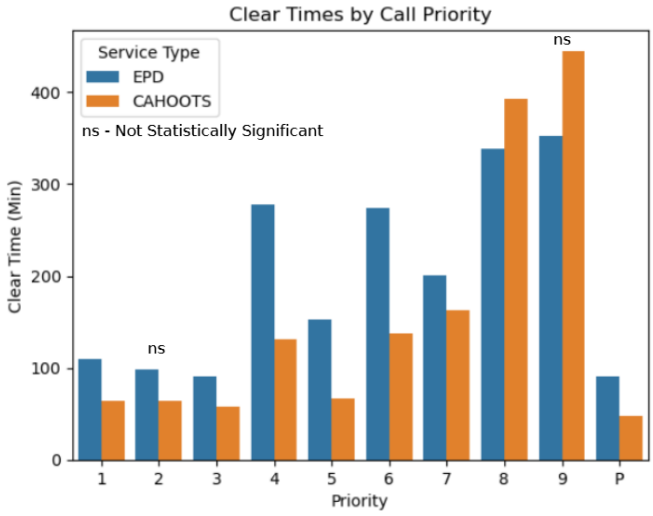
Call natures are where most of the not statistically significant results are for this project. This is likely due to the subjectivity and vagueness of the labeling of call natures. There are 293 different call natures in the data set. For example, the differences between “ASSIST PUBLIC – POLICE” and “PUBLIC ASSIST, CAHOOTS” is difficult to discern because both call natures were responded to by both CAHOOTS and EPD. The graph is sorted by total number of calls responded to (by CAHOOTS) with “ASSIST PUBLIC – POLICE” having the highest number. This means that for the two most common call natures, CAHOOTS does arrive faster than EPD.



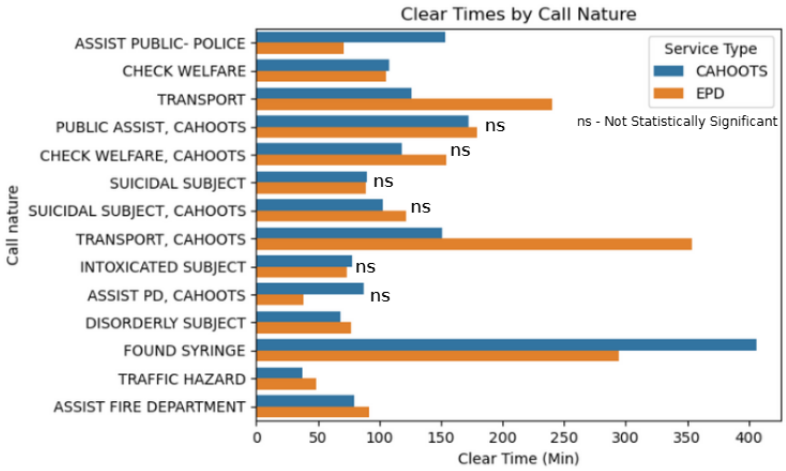
CAHOOTS and EPD are much more similar when it comes to clearing calls by the hour of the day than when arriving to calls by hour of the day. Each is faster than the other for about an equal proportion of the day. CAHOOTS spike in clear time happens a few hours later than the spike that EPD experiences.



CAHOOTS is faster than EPD on Sundays, Mondays, and Tuesdays while EPD is faster on Wednesday, Thursday, Friday, and Saturday. It is interesting that the days on which each is faster are consecutive, with CAHOOTS being faster in the beginning of the week while EPD is faster at the end of the week.



CAHOOTS is only slower for call priority 8 and is faster for priorities 1, 3, 4, 5, 6, 7, and P. There is not enough evidence to say that there is a difference between CAHOOTS and EPD for priorities 2 and 9.



Call natures for clear times have the same issue as with arrival times, often being vague. For the two most common call natures, CAHOOTS is faster than EPD but for “CHECK WELFARE” it is very slight. CAHOOTS is much slower than EPD when it comes to transport call natures.

Discussion

Overall, when it comes to arrival times, CAHOOTS is generally slower than EPD. There are possible explanations for this. CAHOOTS arrival times slow down when call volume increases, indicating a possibility that their resources (maximum of 2 vans at any given time) are stretched too thin, causing people to have to wait longer for service. There might also be systematic differences in how CAHOOTS handles calls from EPD (as seen with the “SELF” call source). One possible reason worth noting, but where this data is unable to give a clear answer, is how the absence of sirens affects CAHOOTS. The CAHOOTS vans are not equipped with sirens so they will get stuck in traffic like any other car. Once on the scene though, CAHOOTS is generally faster at clearing a call than EPD.

A caveat to this report is that it is a preliminary investigation into the results and limited by time and resources. Many of the conclusions and reasonings proposed in this report should be investigated further to confirm any relationship. Just because CAHOOTS is slower when there are more calls in an hour, doesn’t guarantee it's because of a lack of resources. It could also be because of another reason such as the general busyness of the city.

Further investigation should be done into how sirens affect CAHOOTS arrival times, what the exact causes of CAHOOTS being slower to arrive are, and how arrival time might indicate clear times for certain types of calls.

GitHub Link

Lwarrine/CAHOOTS-Call-Data-Analysis

More In-depth Data Cleaning Steps

Description	Input	Output	Method
Import packages numpy and pandas	N/A	N/A	Import (python)
Read in 12 csv files (low_memory = false may be needed for years 17 and 19-25)	12 csv files of the form class_data_20XX.csv	12 pandas data frames	Pd.read_csv() (pandas)
Create a list of variable names for the 12 data frames	12 variable names of data frames	1 list of variable names	[] (pandas)
Concatenate list of data frames into 1 data frame and reset index	1 list of variable names	1 data frame that is all 12 data frames combined	pd.concat (pandas) .reset_index() (pandas)
Save data frame as csv file	Combined data frame	1 csv file	.to_csv() (pandas)
Create a new column called dispatch_time by dividing sec_to_dispatch by 60	Column of combined data frame	New column for combined data frame	Column operations (pandas)
Create a new column called arrival_time by dividing sec_to_arrv by 60	Column of combined data frame	New column for combined data frame	Column operations (pandas)
Create a new column called clear_time by dividing sec_to_clear by 60	Column of combined data frame	New column for combined data frame	Column operations (pandas)
Create a new column called CAHOOT if primeunit is equal to CAHOOT	Column of combined data frame	New column for combined data frame	Column operations (pandas)

Create a new column called hour by taking hour from calltime	Column of combined data frame	New column for combined data frame	pd.DatetimeIndex().hour (pandas)
Create a new column called DoW by taking day of week from calltime	Column of combined data frame	New column for combined data frame	pd.DatetimeIndex().dayofweek (pandas)
Drop any rows where clear_time, arrival_time, or dispatch_time have missing values	Combined data frame	Filtered data frame	.dropna() (pandas)
Drop any rows where clear_time, arrival_time or dispatch_time are less than or equal to 0	Combined data frame	Filtered data frame	Column operations (pandas)
Drop any rows where callsource is equal to "SELF"	Combined data frame	Filtered data frame	Column operations (pandas)

More In-depth Methods Steps

Description	Input	Output	Methods
Function to find the p-value between CAHOOT and non-CAHOOT datasets for different time types (dispatch, arrival, clear)	• Data frame • Column of data frame • Value to compare in column • Time type	Average for each group and p-value between groups	• Independent T-test (scipy.stats) • Column operations (pandas)
Plot arrival times by hour of day between CAHOOTS and non-CAHOOTS calls	Cleaned data frame hour and arrival_time columns	Graph of arrival times by hour of day	Lineplot (Seaborn/Matplotlib)
T-test for arrival times by hour of day for	Cleaned data frame hour and arrival_time columns and array from 0-23	P values for arrival times between CAHOOTS and non-CAHOOTs call for every hour	ttest function (Made earlier)

Plot arrival times by day of the week between CAHOOTS and non-CAHOOTS calls	Cleaned data frame DoW and arrival_time columns	Graph of arrival times by day of week	Lineplot (Seaborn/Matplotlib)
T-test for arrival times by day of week for CAHOOTS and non-CAHOOTS calls	Cleaned data frame DoW and arrival_time columns and array from 0-6	P values for arrival times between CAHOOTS and non-CAHOOTs call for every day of week	ttest function (Made earlier)
Plot arrival times by call priority between CAHOOTS and non-CAHOOTS calls	Cleaned data frame priority and arrival_time columns	Graph of arrival times by call priority	Barplot (Seaborn/Matplotlib)
T-test for arrival times by call priority for CAHOOTS and non-CAHOOTS calls	Cleaned data frame priority and arrival_time columns and array of call priorities	P values for arrival times between CAHOOTS and non-CAHOOTs call for every call priority	ttest function (Made earlier)
Plot arrival times by top 15 call natures between CAHOOTS and non-CAHOOTS calls	Filtered data frame of top 15 call natures. Columns nature and arrival time from that data frame	Graph of arrival times by call nature	Barplot (Seaborn/Matplotlib)
Repeat all steps but with clear_time column instead of arrival_time column			