Ella Cobbs
DSCI 410
Project Report

Research Questions

- What was the change in the percentage of calls handled by other agencies after CAHOOTS added a second van and expanded their service hours in Eugene?
- Has this change impacted the percentage of calls diverted from other agencies to CAHOOTS?

Background

The focus of my research questions is to dive deeper into how much CAHOOTS is used in comparison to other agencies. By specifically focusing on the expansion of their services through the addition of the second van and extension of their hours, a comparison can be made before and after this change to analyze whether it was helpful in allowing CAHOOTS to answer more calls. Also, looking at how many calls are diverted to CAHOOTS shows how they help alleviate some calls from other agencies.

If CAHOOTS is able to answer more calls following this expansion, not only will that be better for the Eugene community, but it will also provide reasoning for further growth. It also allows other agencies, like the Eugene police, to focus on cases more suited to their skill set. This analysis will showcase not only how much CAHOOTS is used but also prove that CAHOOTS should continue to be invested in as it is crucial to the Eugene and Springfield area.

Data

The Eugene CAD data used in this project comes from the Eugene government and is publicly available, but access must be requested. Each row in the dataset represents a separate call and contains information about the date and time, the reason behind it, and who responded to the call. A limitation of the data is that the call signs of who was called first and who responded to the call must be converted to the agencies. However, the dataset contains a lot of observations, each with a lot of information about the call, which is a strength of the data.

Cleaning Steps:

- Downloaded CAD data and loaded in Jupyter Notebook
- Dropped unnecessary columns
    - Only keeping zip code, primary unit call sign, responding unit call sign, and date
- Converted zipcodes from strings to integers
- Renamed columns for clarity and simplicity
    - Call_Zipcode : Zipcode, PrimaryUnitCallSign : Called, RespondingUnitCallSign : Responded
- Kept calls only made in the Eugene area based on zip code
- Converted call signs in Called and Responded to "CAHOOTS" if they matched CAHOOTS call signs
- Dropped time from the "Call_Created_Time" column and converted dates to DateTime in the new "Date" column
- Saved clean data set

Methods

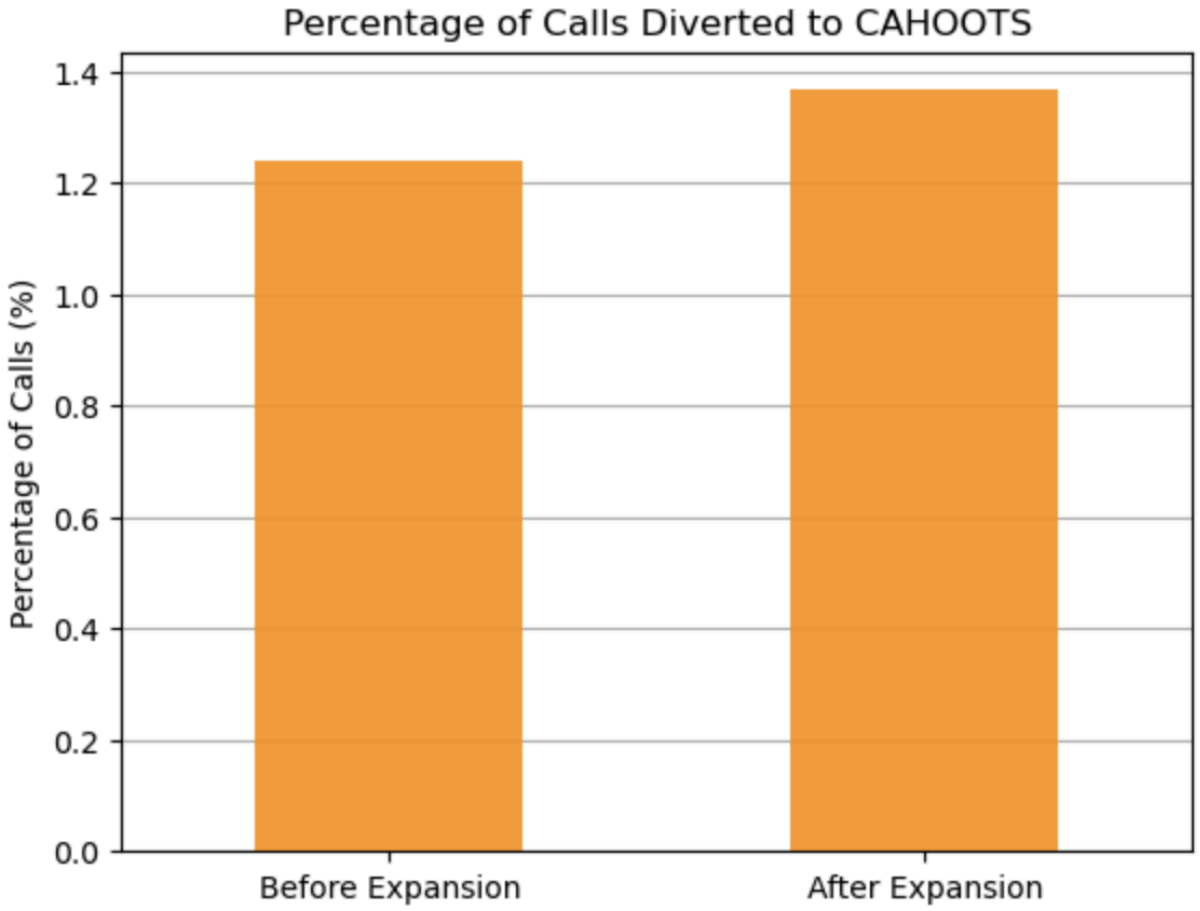
- Defined a function to calculate z-score: input is the portion of total calls for the year before expansion, then the year after, and then the total number of calls for the year before and after -> output is the z-score of the difference of two proportions and p-value of that z-score
    - A z-score can be used to determine if the difference between two proportions is significant. It does this by assuming the null hypothesis to be true, that there is no difference between proportions. It is calculated by taking the difference of proportions and dividing by the standard error. Z-score = (p1 - p2) / sqrt(p(1-p)/n)
    - Where p1 = first proportion, p2 = second proportion, p = total proportion of calls in both samples, n = total number of calls
    - The p-value is then taken of that z-score to ensure the difference is significant, with a 0.05 cut-off. P-value = 2 * cdf (- |z||)
- Calculated z-score for the difference of proportions of the year before and after services were added for calls directly for CAHOOTS and calls diverted to CAHOOTS

- Created a table for the proportion of calls CAHOOTS responded to out of the total number of calls per year
- Used LinearRegression() from sklearn to fit a model to the proportions over the years: input is the years as the x value and the proportion of calls CAHOOTS responded to as the y value -> output is model fitted to the data
    - Linear regression finds a line of best fit that suits your current points so that you can use the equation to make predictions for points outside of your dataset
    - Used this model to predict the proportion of calls over the next 10 years starting in 2025
- Created a bar chart of the percentage of calls diverted to CAHOOTS before and after expansion
- Graphed the line created through linear regression against the real data points
- Created a line graph for the percentage of CAHOOTS calls over all of the data being looked at, with a line added to showcase when the expansion occurred

Scripts & READMEs : https://github.com/ellacobbs/ADS4SJ-Project

Results

Figure 1: Difference in calls diverted to CAHOOTS



Above in Figure 1, it can be observed that the percentage of calls diverted to CAHOOTS is slightly larger following the expansion. The difference between the two is around 0.2%, however the z-score of that difference is -3.15 which is significant on its own, but the p-value of that z-score is 0.0016, further ensuring the significance of the difference between them. For the purpose of this analysis, a diverted CAHOOTS call was when someone else was called, but CAHOOTS responded. These percentages come from the amount of diverted CAHOOTS calls out of all CAD calls for the year, not just CAHOOTS calls.

Figure 2: Percentage of CAHOOTS calls over time

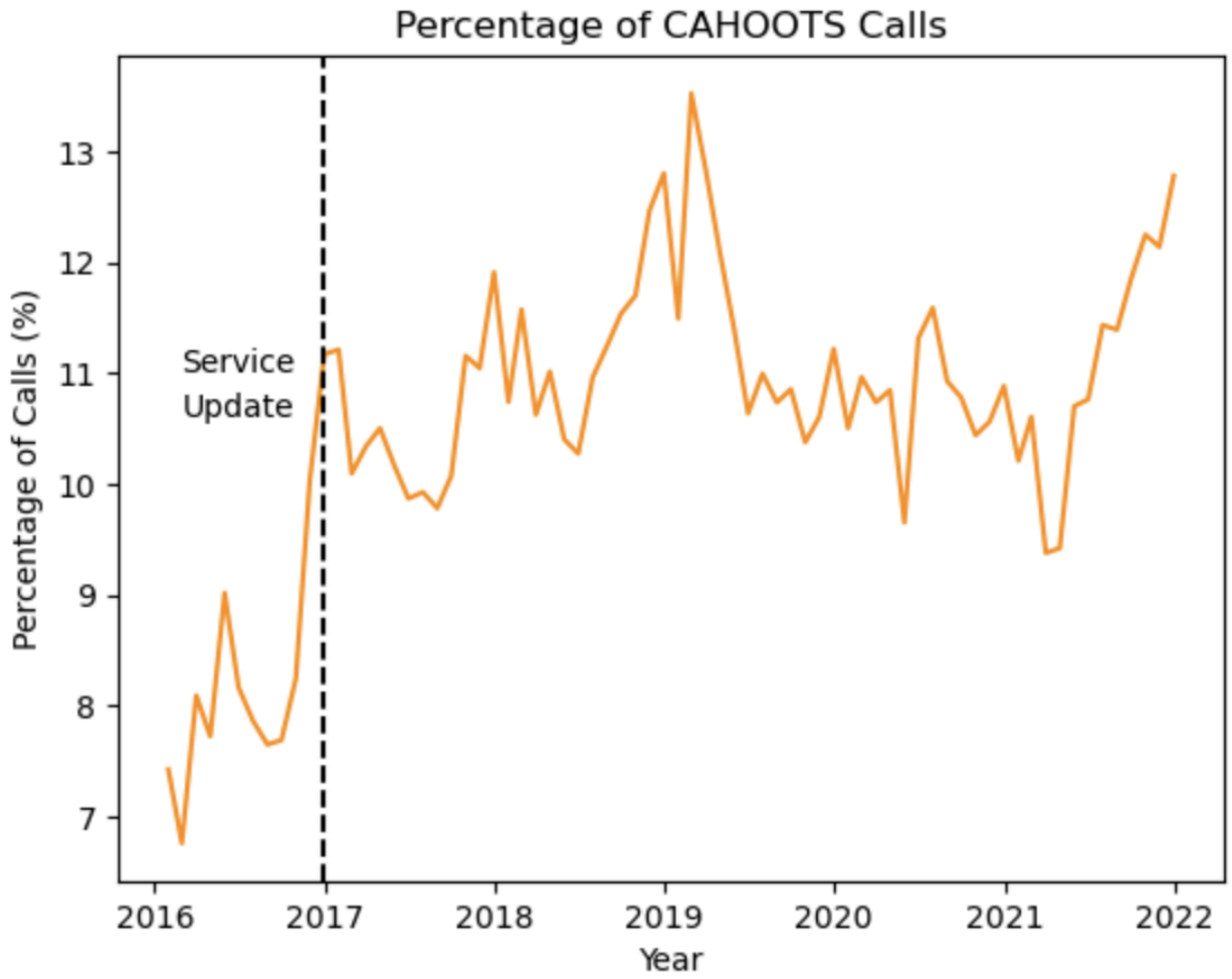
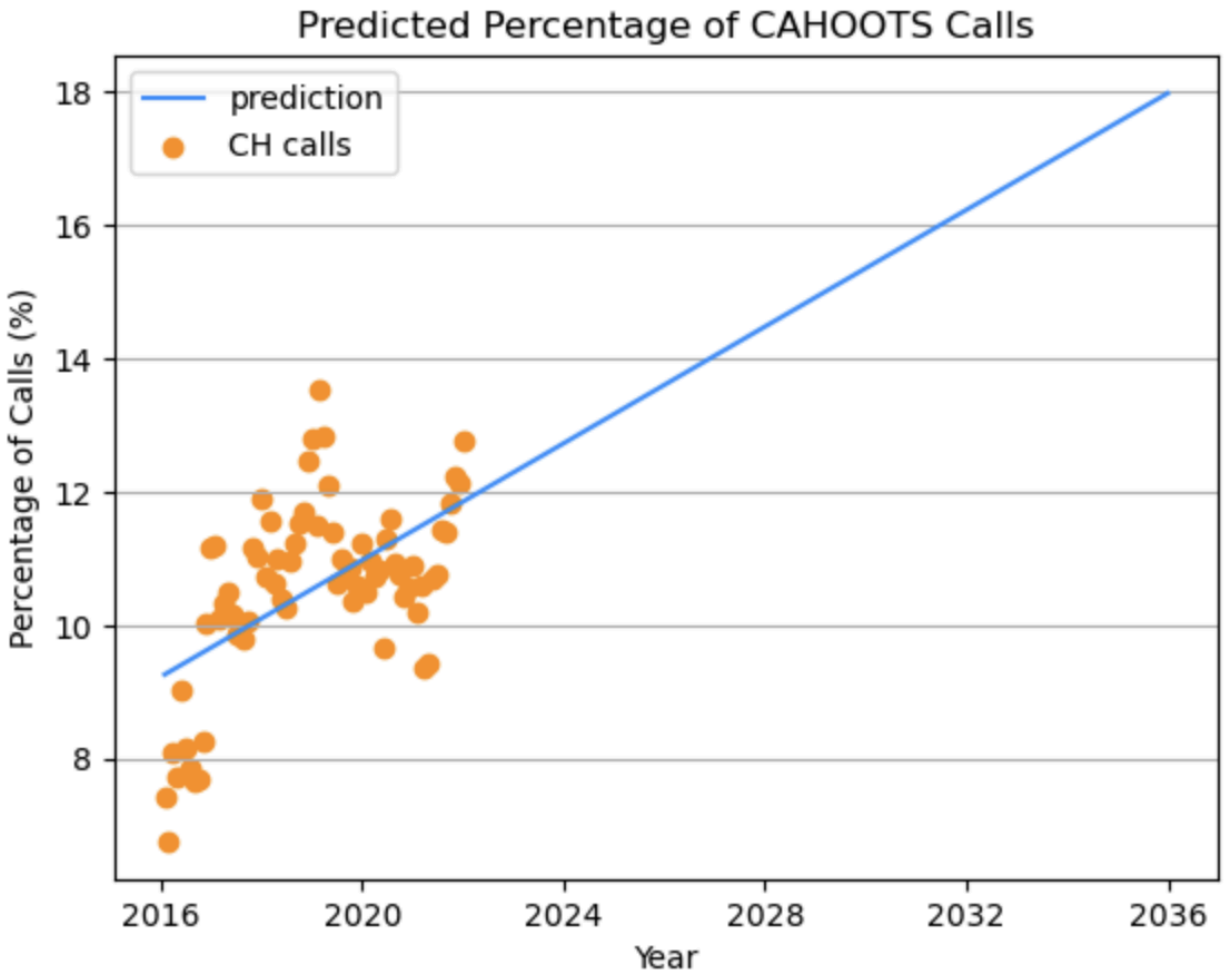


Figure 2 showcases the percentage of calls CAHOOTS handled from 2016-2021, with the percentages being calculated on a monthly basis. These calls represented when CAHOOTS was initially called and responded to the call. A line for when the services were updated was also added for easier comparison. It can be observed that 2016 starts at around 7%, while 2017 starts at 11%. This is quite a jump that happens, and while the percentage dips in the middle of 2017, this looks like a pretty common trend throughout the years and is likely explainable by other factors, like season or weather, for example. There is also a dip in 2020, probably due to the COVID-19 pandemic. However, the percentages peaked in 2019, which could suggest that without the pandemic, current numbers could be even higher.

When comparing the year before the expansion, 2016, with the year after 2017, the difference between the two received a z-score of -21.61, which is incredibly large for a z-score. But still, in

order to ensure the significance, the p-value was taken once more with a value of 1.43e-103. This p-value is impressively small, signifying that the difference between 2016 and 2017 is quite significant.

Figure 3: Linear regression predictions of the percentage of CAHOOTS calls



The final graph shown in Figure 3 illustrates how the CAHOOTS percentage of calls they handle is expected to grow. The predictions made through linear regression are graphed with the blue line with the actual data points presented in orange. As seen in Figure 2 as well, the growth of CAHOOTS calls does not follow an exactly linear pattern, but this prediction tends to go through what about the average percentage throughout the year would be. It should reach around 18% by 2036 based on this line, but realistically, there is a limit on the amount of growth available. This

is because CAHOOTS can only handle a certain number of calls given their resources and are also only defined to answer a specific subset of calls that come into 911.

Discussion

Based on the analysis performed comparing the number of calls handled before and after the expansion, the increase in the percentage of calls answered is likely due to the expansion that occurred at the beginning of 2017. While direct causation can not be conferred, there is a strong correlation between the expansion being followed by more calls served. This is because the p-values for both diverted and full CAHOOTS calls were both well under the 0.05 cut-off. Because there is also a prediction for a continuation of growth as a result of the linear regression, it would be strongly encouraged for CAHOOTS to continue extending its resources in order to be able to answer the new demand of calls proficiently.

The main strength of this analysis was how low both p-values were. Because they were much lower than they needed to be, under 0.05, the answers to the questions were more confident in the recommendations' accuracy.

It would be interesting to look further into what agencies CAHOOTS is taking calls from. Unfortunately, with this data set, looking at the other individual agencies, police, fire, and medical, was not possible, but I think investigating whether all calls diverted to CAHOOTS are from police would be important. So, a weakness of this analysis was that this data was unable to confirm which agency or agencies have their percentage of calls handled shrinking as CAHOOTS' grows.

To return to the questions, the difference in the percentage of calls handled by CAHOOTS following the addition of the second van and extension of service hours was around 3.79%, and it was found to be significant. The difference between diverted calls was around 0.2%, and it was also significant.

In order to help further prove the causation of the expansion causing an increase in calls, further research could look deeper into possible confounding variables that may have also caused this change. Some examples of this could be weather, specifically extreme events like heat waves or freezing temperatures, economics or the job market, and average house prices compared with the average salary. I chose these variables specifically because they may directly impact the number of people who are in need of CAHOOTS' services.