Optimizing Policing for Santa Clara/Ivins Police Department

Matthew Gergley[1], Ewan Townshend[2], and Cameron Netherton[3]

[1]Department of Mathematics, Utah Tech University
[2]Mentored by Dr. Vinodh Chellamuthu
[3]Industry Liasons: Captain Rodgers & Lieutenant Studley

April 2023

Contents

1 Abstract

When there is a 911 call made, chances are a police officer will be sent to the location of the call along with other required emergency services. In a big city, this yields little issue since there is normally a sufficient amount of officers to handle a volume of 911 calls, but what about a smaller city? The goal of this project is to provide an answer to the SCIPD (Santa Clara/Ivins Police Department) of where they should be allocating officers and when exactly throughout the day and much more. The key problem area is that there aren't many officers to go around thus making this a critical optimization problem. Due to the randomness of 911 calls, we explore aspects of probability and data analysis through a given data set of 911 calls from 2017-2022 in Santa Clara and Ivins, Utah. The outcome of our results will best allocate SCIPD's sources and minimize average response time and workload per officer to allow for the best active and passive policing executable.

2 Introduction

A city's emergency service, and more specifically, police departments response time and efficiency is key to maintaining order and civility in a city. And especially important if the numbers of a specified workforce is minuet. This is the scenario in Santa Clara and Ivins, Utah where the SCIPD is made up of currently 15 officers in a jurisdiction of roughly 20,000 citizens. A worthy note is that an average citizen to officer ratio is 2 officers per 1,000 citizens and furthermore in Utah it is suggested 1.5 officers per 1,000 citizens. In an ideal world, a department would have whatever resources they need but that's not always the case of course as we see here. So working with limited sources is what makes it necessary to exploring how to optimize policing with the given number of officers. Another thing to note about this particular department is that currently patrol officers follow shifts 4am-4pm and 4pm-4am, which will be called the "Day" and "Night" shift respectively.

This paper addresses the following research questions proposed by Santa Clara/Ivins Police Department:

- Where are the best places for officers to be?
- What times should they be in those areas?
- What time has the most occurrences of phone calls?
- Amount of calls per service per officer?
- Where should passive policing occur?
- How should active policing be conducted? Is there a patrol route that should be followed?

The first step, as in most real-world problems, is to look into the data available to find trends and extract key elements that can help us reach our end conclusions. For the Santa Clara/Ivins Police Department, they receive roughly 4,000 (911) calls per year. This is smaller compared to most departments around the US such as the New York Police Department (NYPD) which might receive around 4,000 calls in just one day (NYPD, 2023). Now in a mathematical analysis, having the call data as all words and descriptions with address, call nature, date, et al. is not the most useful obviously. Thus translating words to numbers is a key component and step that this model will explore.

Models do exist for optimizing patrol routes and even garbage truck routes which in essence follow a similar structure of integer programming. This model and solution will take a different approach while absorbing some key factors such models. The key to remember is that response time is the number one priority when talking about emergency services as well as balancing the an emergency workers stress and workload as they endure strenuous moments in a given shift. How can we take descriptive data and achieve our previously stated research questions?

3 Data Analysis and Solutions

3.1 Preliminary Data Analysis

The files sent to us by the Santa Clara and Ivins Police Department contained data dating from January 1st of 2017 to the 18th of August 2022. The original data set contained data from all the emergency services (Fire, Police and EMS), due to this all calls where the police were not involved were removed. Each of the CSV files were then converted into a full data set. The data set itself was extremely clean, with very few null cells/cells only containing blank spaces:

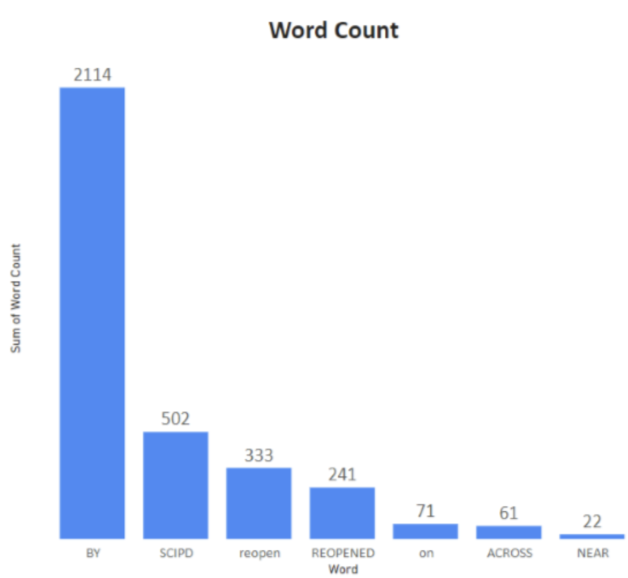
Total Cells in Data Frame:	$293,381$
Cells that are blank:	0
Cells that are a single space:	3
Cells that are a double space:	0
Cells that are a triple space:	0
Cells that have 15 spaces	1
Accuracy:	99.999%

Only one of these cells which is empty came from a column that was used in the general analysis, and so this row was removed. This left us with a total data set of 26,370 rows spread across 5 years:

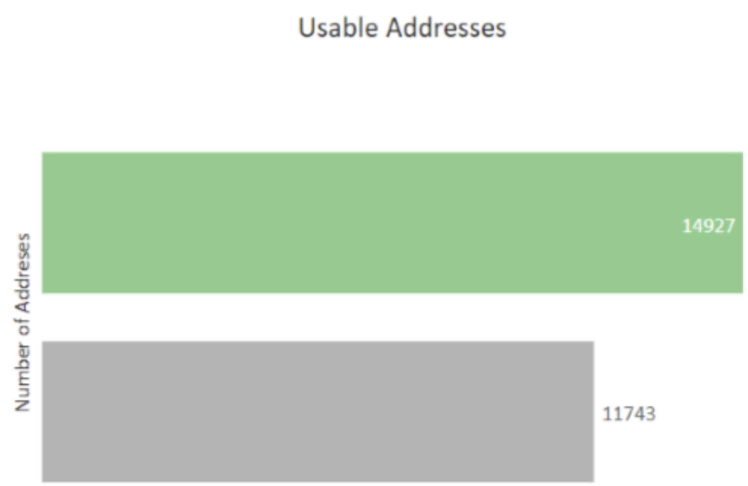
Calls per Year

Year	Number of Calls
2017	4275
2018	4665
2019	4746
2020	5116
2021	4901
2022	2967
Total	**26670**

There was a significant drop off in calls for 2022 due not having data on calls placed after August 18th, 2022.The only area in the data set which required manipulation before use was the addresses, as the operators often included keywords such as 'near' or 'by' for calls where the address was not as simple as a domestic residence/commercial business. We found that in the data set the words 'By' and 'SCIPD' were used most often (outside of the terms North, East, South or West or anything that could be in a street name):



Once the keywords were identified, we removed these words from the address column in the data set to maximize the number of usable addresses:



The total number of addresses used in our models was 14,927, roughly 56% of all the addresses given. These addresses were then converted into coordinate form (longitude and latitude) and used in calculations of both the active policing spots and heat maps.
As well as manipulating the addresses, several calculated columns were created to help with the analysis. These include an hour column for the 'Average number of calls per hour' visual and a seasonality column, which uses December to February as winter, March to May as spring, June to August as summer and September to November as Fall. This field was then able to be used as a filter in many visuals to test whether seasonality affected the analyses. The most significant column created was the 'Call Severity Weight' column. Not all calls should be given equal weighting, as some calls (such as murder or robbery) are more urgent and require more resources than others (such as a routine check or a noise complaint). This can be likened to the idea that if person 1 was given 5 $1 bills and person 2 was given 5 $20 bills, by count of bills they have the same number. However, each $20 bill is worth more than each $1 bill, and so person 2 has more money. 5 critical calls in 1 night would constitute a much heavier workload than 5 checkups and this weighting system intends to show this. One of the columns in the original data set was the 'call nature' column, indicating why 911 was called. The Santa Clara and Ivins Police Department gave us a list of the main call natures, and

we adapted this list to fit into the larger list of call natures in the data set. Each call nature was given a ranking between 1 and 5, with 1 being the most severe calls and 5 being the least severe calls. We then took the reciprocal of the score and created the call weight. This can be seen in the table below:

Call Nature	**Score**	**Call Weight**
Agency Assist	5	$\frac{1}{5} = 0.2$
Chest Pain-10	1	$\frac{1}{1} = 1$
Animal Problem	3	$\frac{1}{3} = 0.33$
Family Fight	2	$\frac{1}{2} = 0.5$
Animal Problem	3	$\frac{1}{3} = 0.33$
⋮	⋮	⋮

Rather than using a count of all the calls for our calculations, the sum of the call weight was used in all our visuals.

The questions that we intended to answer ranged from general analysis, such as what the busiest day of the week is and busiest day of the year, to more in depth analysis to determine the best time to switch shifts and the best places for active policing using the coordinates from the addresses. The best time to switch shifts depends on what shift times see a 50/50 split in weighted calls between shift one and shift two and what would be best for the personal lives of the officers themselves. A shift change time of noon and midnight would be tough on the officers and would damage morale. Equally, if shift one takes all the calls this would exhaust the officers in shift one.

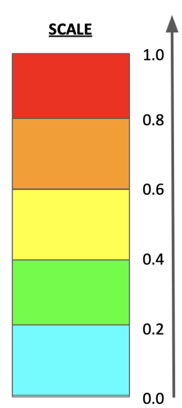
The solution to this problem was attained by creating a new calculated table in Power Bi. This table has columns for the start of shift one and the end of shift one. The percentage of calls in shift one was created by using the 'calculate' function and adding filters to indicate only some call weights that are after the start of shift one and before the end of shift one. This number was then divided by the total sum of calls to get the percentage of calls in shift one. The percentage of calls in shift 2 was then calculated by doing one minus the percentage of calls in shift one:

Shift One Start Time	Shift One End Time	Calls In Shift One	Calls In Shift Two	Calls In Shift One %	Calls In Shift Two %
00:00:00	12:00:00	3593	6144	36.90%	63.10%
00:30:00	12:30:00	3723	6013	38.24%	61.76%
01:00:00	13:00:00	3881	5856	39.86%	60.14%
01:30:00	13:30:00	4047	5690	41.57%	58.43%
02:00:00	14:00:00	4230	5507	43.44%	56.56%
02:30:00	14:30:00	4443	5294	45.63%	54.37%
03:00:00	15:00:00	4651	5086	47.77%	52.23%
03:30:00	15:30:00	4897	4840	50.29%	49.71%
04:00:00	16:00:00	5108	4629	52.46%	47.54%
04:30:00	16:30:00	5351	4386	54.96%	45.04%
05:00:00	17:00:00	5595	4142	57.46%	42.54%
05:30:00	17:30:00	5773	3964	59.29%	40.71%

And as we can see the shift times that result closest to a 50/50 split in call volume is having shift one be 3:30am-3:30pm and shift two be 3:30pm-3:30am.

3.2 Heat Maps

To get a general idea of the spread of calls we can use the now numerical representation of call location, longitude and latitude, to develop a heat map of the 911 calls which serves as a histogram of calls on a physical map. Also, it has the ability to be restricted by season, shift times, and a given year thus allowing to be able to compare amongst different times over the years to see if there are trends worth noting. Another very useful part of the heat maps is that it allows us to compare our numerical results to a graphical representation to see how they compare and if we can derive the same conclusion from both. How the call weights play into this can be seen from this color chart based on call weight:



The development of the heat maps was done via Python and the geopy and folium packages. The idea was taking all of the address and converting them into longitude and latitudinal coordinates for plotting on a map and also so we have numerical representations for further analysis. Then for example, if a call at a certain coordinate location had a call nature of "Animal Problem", we can see from the table above in **3.1** that it has a call weight of 0.33 so at that coordinate there would be a circle of the lime green color. And let's say that there again is another call of the same nature at the same location, the score accumulates to 0.66 and changes the color at that coordinate to orange and once the call weight at a given coordinate is 1.0 or greater it is red no matter the additional call weights that might occur. So using the call weighting system we were able to develop heat maps from 2017 to 2022 divided all the way down to what season and shift time resulting in a large amount of images but we can highlight some of the maps below:

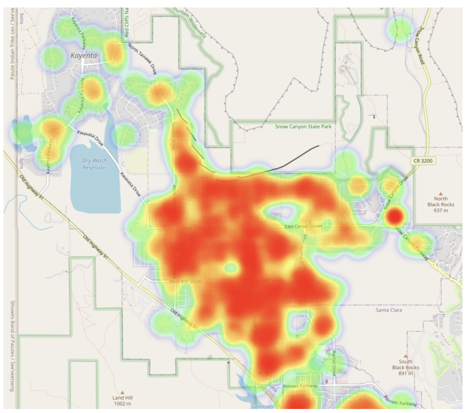


Figure 1: Ivins 2017

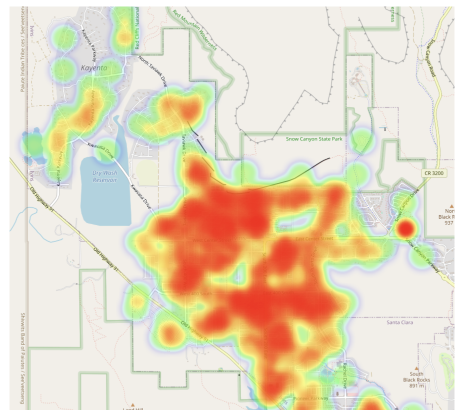


Figure 2: Ivins 2021

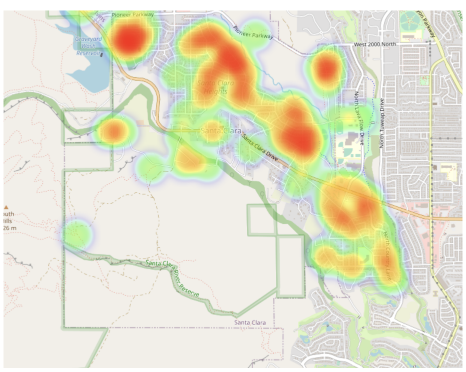


Figure 3: Santa Clara 2017

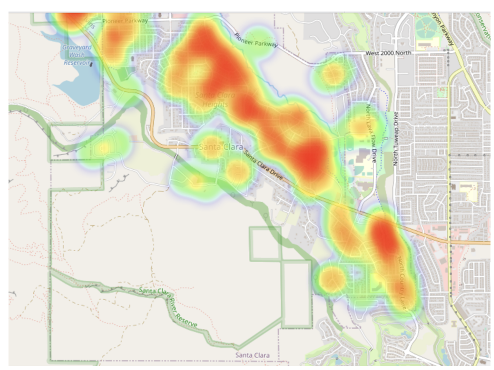


Figure 4: Santa Clara 2021

From the following images there are not much noticeable changes but as we can see there is an increase in volume of the amount of 911 calls which is justified since the population in Ivins and Santa Clara have been steadily increasing over the years. Now looking at 2017 maps comparing opposite seasons; summer and winter, spring and fall, during the day shift we are able to see some important changes:

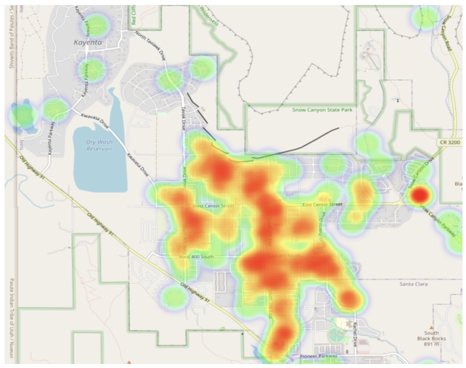


Figure 5: Ivins Spring

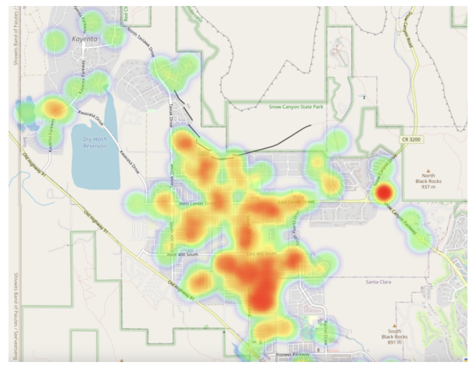


Figure 6: Ivins Fall

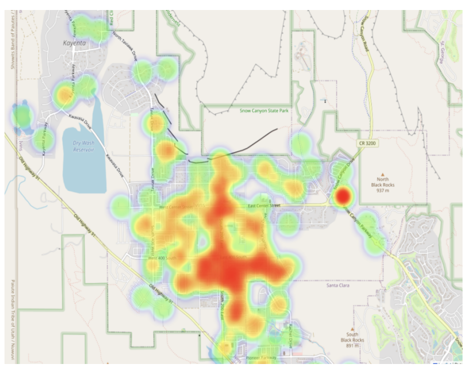


Figure 7: Ivins Summer

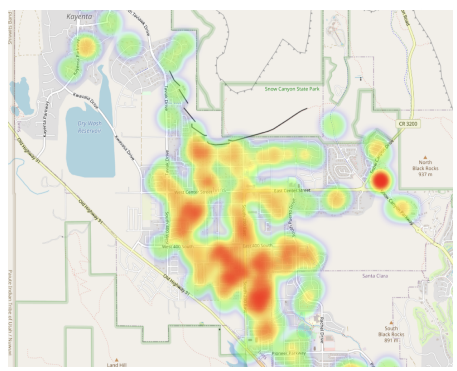


Figure 8: Ivins Winter

In Ivins we see that the warmer season, namely spring and summer, we see a bigger spread in calls as well as more volume of calls as well which makes sense since it is warmer and people are more active which results in more possibilities of a 911 call needing to be made. However if we look at the same analysis for Santa Clara we see something different:

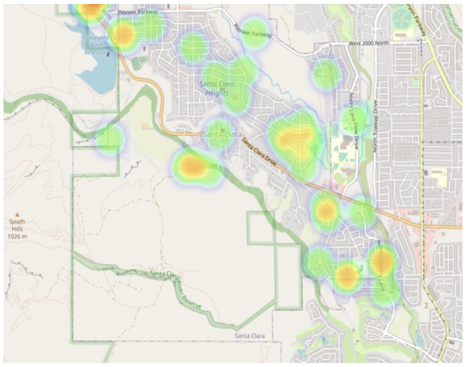


Figure 9: Santa Clara Spring

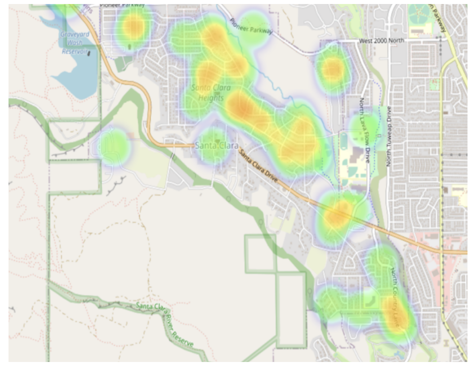


Figure 10: Santa Clara Fall

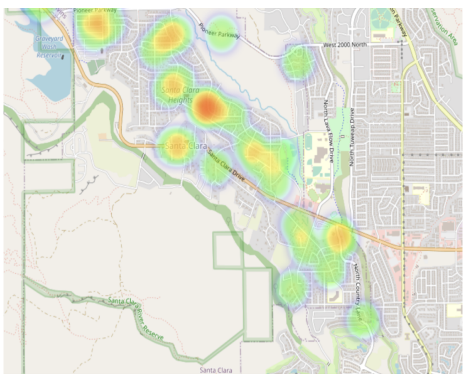


Figure 11: Santa Clara Summer

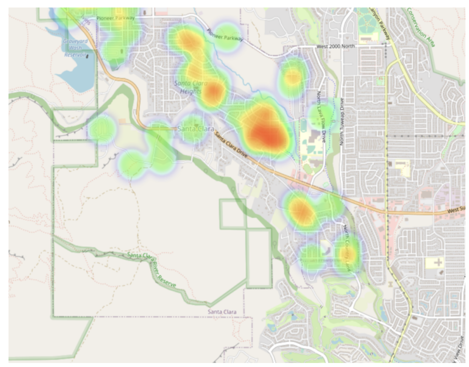


Figure 12: Santa Clara Winter

As we see in Santa Clara it actually follows the opposite case of Ivins and of what is expected. There appears to be a greater volume of calls in Santa Clara in the colder season, namely fall and winter and a bigger spread of calls in those seasons as well. This ends up being supported by data analysis which will be highlighted later on in this paper.

3.3 Centroids

The objective of the coordinate analysis is to see if there is an optimal patrol route that can be taken by the SCIPD to minimize the distance that needs to be covered to get to a call. This would minimize response time and maximize the utility of the officers. The solution to this issue was to use a weighted average for the coordinates. For each call placed, you can take the longitude and latitude and multiply them by the weights. You can then separate the longitude and latitude for each call and create a longitude list and a latitude list. For each list, you sum all the coordinates multiplied by the weight and divide that by the sum of all weights. (Note: x_i denotes the longitudinal coordinate,y_i represents the latitudinal coordinate, and w_i denotes the call weight for the corresponding coordinate location). Below is the linear combination that results:

$$\frac{1}{w_0+w_1+w_2+\cdots+w_n} \cdot \left[w_0 \cdot \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + w_1 \cdot \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + w_2 \cdot \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} + \cdots + w_n \cdot \begin{bmatrix} x_n \\ y_n \end{bmatrix} \right]$$

Thus, using this linear combination method due to the nature of having call weights being necessary, we are able to apply this as previously noted to any sort of subset of the data. Thus what will be important in our analysis is to find passive policing locations (where officers should park their cars) and active policing locations (how officers should move through the city(s)) which will be important for developing a patrol route. To find the passive policing locations, we can implement the average of weighted linear combinations for Santa Clara and Ivins to get each cities best passive policing locations which will allow for the fastest average response time. Then, we can split both Santa Clara and Ivins into quadrants to find active policing locations, by finding each quadrantal centroid using the same average of the weighted linear combination method as before just splicing the data for the given quadrant and city. Again, we further split it all the way down to seasons and shift times being Fall, Winter, Spring, and Summer and the day shift being defined as 4am-4pm and night shift being 4pm-4am. Hence we can see if seasonality or a specific shift requires different policing locations to optimize response time which would result in a different patrol route as well. Below is the findings with the large circles being the passive policing locations and the smaller circles surrounding are the quadrantal centroid which are the active policing locations (Note: Santa Clara is denoted as the dark blue color and Ivins is denoted as the light blue):

Performing additional analysis into seasonality and day and night filters, there was not much difference in the overall locations however it is fascinating and worth noting some of the differences that might seem minimal but the difference can make a response time to a 911 call that much quicker and potentially save a life. So below we have two images of results in Santa Clara:

In the top image we see that the passive policing spot (orange circle) in Santa Clara moves a block southeast depending on if it is day or night. And in the bottom image we see that in three of the four season, an active policing spot is in one location, but in another season, spring, it is a block over to the east.

3.4 Patrol Route

Using the quadrantal centroids, we then are able to derive a proposed patrol route. Since we have 4 active policing coordinates in both Ivins and Santa Clara, we can use our knowledge of permutations to find which way the officers should travel through the prescribed coordinates. In our case, the classic permutation formula,

$$nPr = \frac{n!}{(n-r)!}, \tag{1}$$

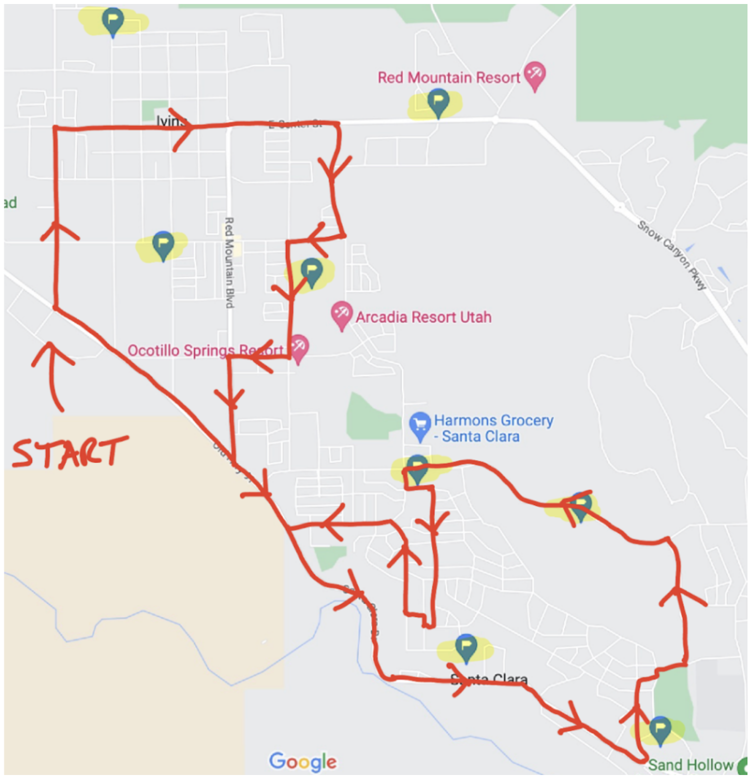
which boils down to simply **n!**, since we are choosing all the coordinates for each permutation. Thus there ends up being 4! or 24 different ways one can travel through each cities coordinates thus we want the one that minimizes the distance traveled the most while keeping the officers close to the coordinates so they can have the fastest response time when on the move. So to find the distance of all of these resulting permutations we can perform iterations of the Haversine formula to compute the distance for each permutation which is necessary since we are dealing with locations on a sphere and that sphere being the Earth. The formula to calculate the spherical distance between longitudinal and latitudinal locations is defined as,

$$d = 2r \arcsin\left[\sqrt{\sin^2(\frac{\phi_2 - \phi_1}{2}) + \cos(\phi_1) \cdot \cos(\phi_2) \cdot \sin^2\frac{\lambda_2 - \lambda_1}{2}}\right], \tag{2}$$

where ϕ_i are the latitudinal part of the starting and ending coordinates and λ_i are the longitudinal part of the starting and ending coordinates, respectively. Also r is the radius of the sphere which in our case is the Earth and the result d is the distance between the coordinates. And thus using the resulting permutations we can sum up the distances between the given permutation. Hence our total distance for each permutation can be denoted as,

$$D_i = \sum_{i=0}^{n} d_i \Rightarrow min\left[D_i\right], \tag{3}$$

and in our instance $n = 4$ since we are testing permutations of 4 active policing coordinates and D_i is the total distance for the given permutation which we have 24 for each city as previously stated. Iterating this using Python, we can obtain the permutation with the shortest total distance between all points. Note, this is absolute distance and does not take into account the need to travel on a physical road however after testing the routes via road they followed the same results and order as absolute distances. The rationale behind separating Ivins and Santa Clara separately instead of doing a permutation with all 8 coordinates is because the quickest way between Ivins and Santa Clara is well defined so we are mainly concerned with patrolling through each respective city. The proposed patrol route is as follows where officers would travel between Ivins and Santa Clara via Old Highway 91:



In this route, the active policing centroids are highlighted with a yellow color and also there is an arbitrary start point simply to define the routes direction however it can of course be started at any point. And also note that it is not a necessity to have the route physically extend to each point but rather maintain a tighter spread to stay in that centralized location while being able to bring officers to roads with easy transitions to the hot spots. This route does however bring officers close to the quadrantal centroids which is a plus. And of course, this route again, gives the officer the minimum distance traveled and will result a lower average response time. The minimum distance is also important since it will save the department money on gas usage and also mileage on patrol cars resulting in less maintenance long-term.

4 Conclusion

Our findings have been derived by using mathematics and data analysis and here we can sum up all of our findings. Seasonality is an important component to monitor when considering police scheduling and distributions of calls as seen in the heat maps. Also noting the fact that Ivins receives a majority of the 911 calls and the more severe calls as well. Through data analysis we were able to discover that the optimal shifts is 3:30am-3:30pm for the day shift and 3:30pm-3:30am for the night shift which is a half hour earlier than the schedule they currently follow. Using the centroids, if an officer wanted to perform passive policing then the best spot to park their patrol car is at either the Santa Clara or Ivins Elementary schools as those are the overall centroids in each city allowing the officer to be in a neutral location to have a good response time to any 911 call. Then, using the quandrantal coordinates (for each city, active policing) and permutations we were able to develop the optimal patrol route that brings officers close to areas where calls tend to be so they can have a minimum response time while also travelling less distance which saves money on gas and car maintenance as well.

5 Future Works

Research into this overall department and optimizing the policing with the limited resources is of course not done and has many things that can be explored for further results that can be useful for the department. Adding citation data would be a good next step to further deepen the analysis. Also looking into workloads per officer in the department would be something that could be useful and correlates to when to change shift times as currently it is based off the 50% mark of weighted call volumes which might not actually represent how productive a given officer is.

6 Contributions

- Ewan Townshend: Data and Solution Approach, Poster, Dashboard
- Matthew Gergley: Abstract, Introduction, Heat Maps, Patrol Route, Conclusion, Future Works
- Cameron Netherton: Centroids

Sources used: Chandra, 2019, Thabet et al., 2023, Wikipedia, 2021, NYPD, 2023

7 References

Chandra, V. (2019). Routing optimization for garbage collection. https://medium.com/analytics-vidhya/routing-optimization-for-garbage-collection-495225a67daf

NYPD. (2023). Police department data. https://data.cityofnewyork.us/Public-Safety/NYPD-Calls-for-Service-Year-to-Date-/n2zq-pubd

Thabet, M., Messaadia, M., & Al-Khalifa, A. (2023). Police patrol routes optimization: A literature review. https://ieeexplore.ieee.org/document/10050899

Wikipedia. (2021). Haversine formula. https://en.wikipedia.org/wiki/Haversine_formula