Data Driven Optimization for Police Schedule Design

A. Goulding, R. Morgan, A. Tipper, M. Zheng

April 25 2022

Acknowledgements: Dr. Vinodh Chellamuthu, Captain Micheal Giles, Justin Grenier

Abstract

The Saint George Police Department works tirelessly to enforce the law, and protect its community. On a day to day basis, they receive numerous 911 calls which are tended to based on their priority. With staffing becoming an issue in every workforce, this can make things difficult when attempting to maintain quality of service and being efficient in responding to the calls. Although the Saint George Police Department has been well supported with the difficulty of staffing, it is still paramount to look at alternatives ways to utilize current staff to ensure maximal output. Hence, working to alter shift schedules for the police department is important not only for efficiency but also economically. We developed a mathematical model to predict future trends of incidents within the community. With this information, we gained a better understanding to organize and give weight to numerous statistical categories which we then used to determine the labor needed on a given day. Using our results, we intend to provide the police department with the data so they can optimize their resources.

1 Introduction

1.1 Problem Statement

The demands of the police force are becoming increasingly more challenging resulting in their resources being extremely stretched. The police department is of paramount priority in any community, ensuring all citizens feel safe in their area by the enforcement of laws. However, with the increasing workload for officers and the constant battle with lack of staffing, their ability to enforce the laws effectively is influenced dramatically. Covering 77 square miles, and serving almost 80,000 civilians, the Saint George Police Department (SGPD) has stressed an urge to combat this dilemma especially with the Saint George population growing rapidly. Although SGPD receives great amount of assistance when it comes down to staffing, funding, and other resources, the workload for police officers is immense, making it hard to efficiently respond to the volume of tasks. Additionally, the SGPD voiced that their data was difficult to analyze. Hence, being so disorganized, they cannot effectively use the data such as for spotting trends.

When given the dilemma of lack of staffing, a first instinct might be to hire more staff. However this is not necessarily the best approach, one reason being it may not be financially viable. In addition to this, qualified staff is hard to come by, especially in the police force with there being a decline in interest for this occupation over the recent years. While we cannot control the staffing levels, we can look at alternative ways of distributing current police officers. By analyzing incident data in Saint George over the recent years, we have put together a system that determines what days of the week are predicted to have the highest workload, with workload not only including the number of incidents, but also considering the priority and time that is required for each incident. With this, the SGPD will be able to utilize and schedule their staff in the most optimal way to serve the community.

1.2 Goals

In order to determine how to reorganize the officer schedule, we first need to be able to analyze the volume of work each call takes. Our overarching goal is to develop a new work schedule for SGPD that optimizes their current staffing levels. To do this, we set out to find a method to prioritize each call and be able to weigh them. Thereupon we can use this data to find out which days of the week require more work based on the average weights per day and the SGPD can better determine which days will require more staffing.

1.3 Data/Resources

The SGPD allowed us access to the 911 call data from January 2017 to December 2021. This data contained the time and date of the call, the nature of the call (i.e. what was being reported), the offenses that were later recorded, which departments arrived to the incident (law department, fire department, or EMS department), the incident address, and the progress code given for the call. In addition to this data, we also received all of the accidents and citations that the law department responded to from 2016 through 2021.

2 Initial Trends

When first receiving the data sets from the police department our information was unorganized and spread out through multiple files. Our first objective was to grasp an understanding of how frequently incidents are being reported. Using python, we ran through the sets of data organizing them so that we could properly use the information at hand. Sorting through the data we were able to create a figure which showed the number of incidents reported through 911 calls in the previous six years (Figure 1). This figure gives us a good overview of how many incidents we will be handling and showing the trends of incidents over the most recent time period.

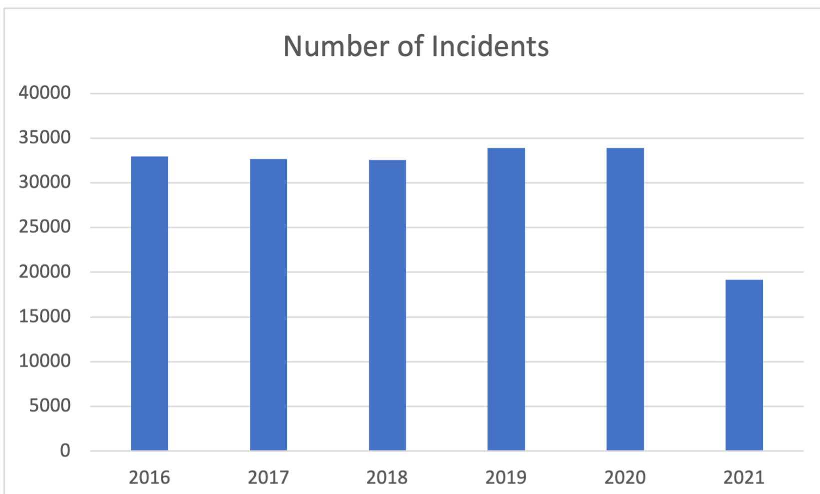

Figure 1: Number of Incidents over the last 5 years

Examining figure 1, we are able to see that from 2016 to 2018 there were similar totals in 911 calls over that time span. While in 2019 and 2020 there was an increase in calls made. 2021, is extremely low compared to the other years. Note that this is due to the fact that we only have data for 6 months of 2021, thus leading to the lower total in 911 calls made in 2021.

Next, we wanted to pull apart our data to show seasonal breakdowns in each year. This is represented in Figure 2. The chart shows that the winter season has the fewest incidents reported while the spring season tends to show an increase in the number of incidents. From this information we look to find trends based on temperature that could show a correlation between more 911 calls with warmer weather. This would then help us understand the amount of field work required dependent on the time of year or weather in a particular time frame.

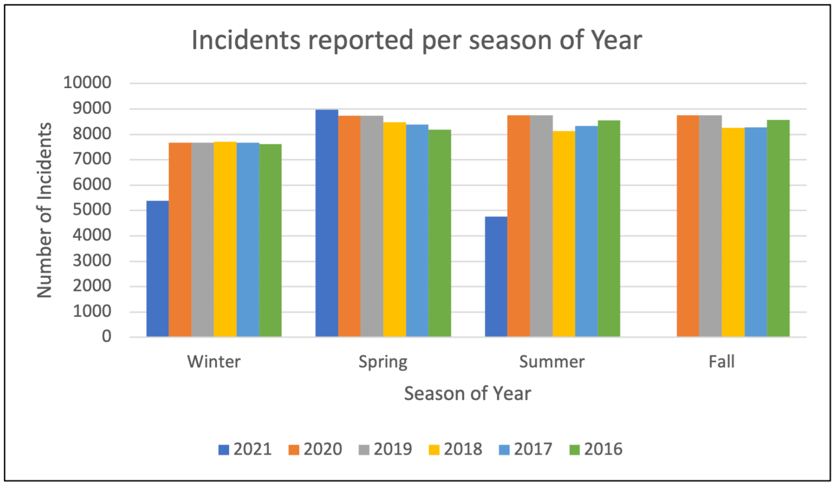

Figure 2: Season breakdown of Incidents

Additionally we broke down the monthly trends in Figure 3. In this figure we are able to notice similar trends as seen in the Figure 2 (seasonal). We found that the winter months of November through February have the fewest number of incidents reported, while the months of May to October seem to have similar totals in reported incidents which are higher in comparison to other months. Again, note that we have data for the first six months of 2021, hence the drop off for the line representing 2021 in Figure 3.

Zooming our scale into a more relevant time frame for our main goal of creating an optimal work schedule, we looked at how days of the week can have an impact on when incidents are reported (see Figure 4). We observed that Wednesdays and Fridays show the highest number of incidents, while Sundays tend to have the least number of incidents. This information, when paired with the priority of each incident happening on each day of the week, can help us understand on a weekly basis which days require less or more manpower, in hopes that this helps the department be better prepared for what may occur during a week.

Our last break down focused on the time of the day each call is being reported at as shown in figure 5. Breaking down the total number of incidents by hour from 2018 to 2021, we notice a similar trend each year. Peaks in reported incidents are found from 8 am to 6 pm. While we also notice lows in the data from 9 pm to 6 am. This will help us in understanding when each shift should start and end as well as times during the day in which additional support may be needed.

These figures were able to give us useful background information on chrono-

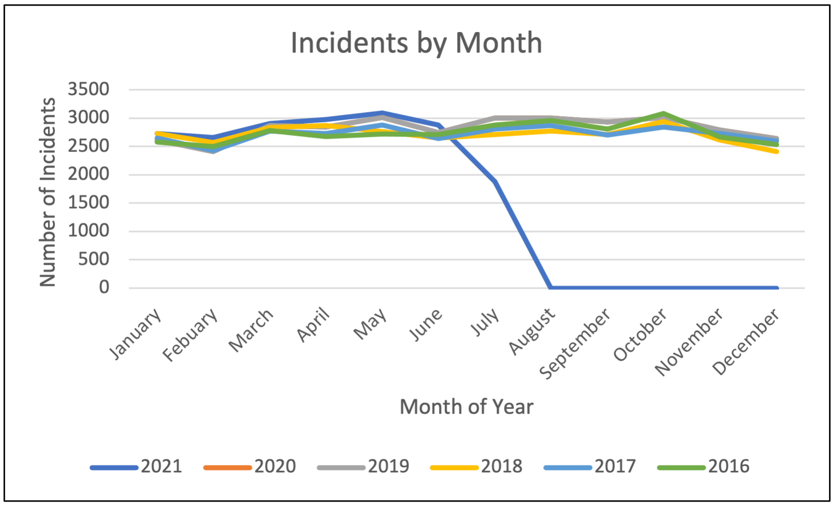

Figure 3: Incident Breakdown by Month

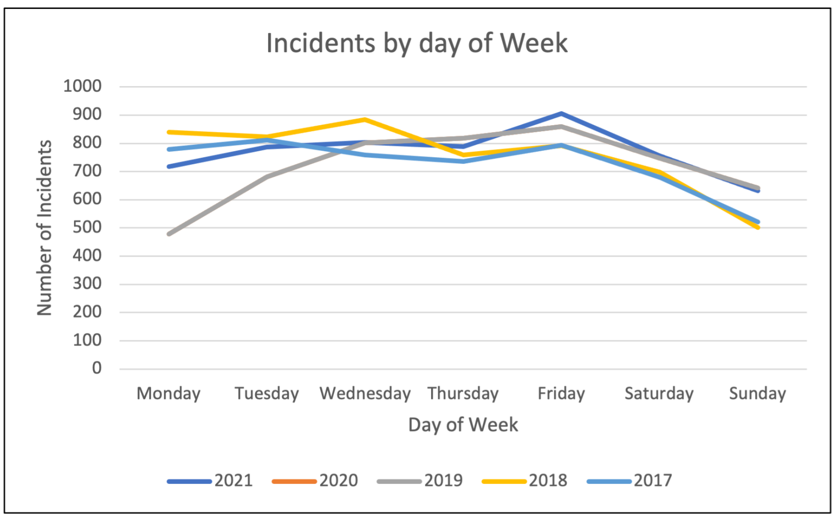

Figure 4: Incident Breakdown by Week

logical trends that can effect the resources needed at a given time. The next step is to pair this information with the priority of incidents to action and to find the trends of overall workload throughout a day and week.

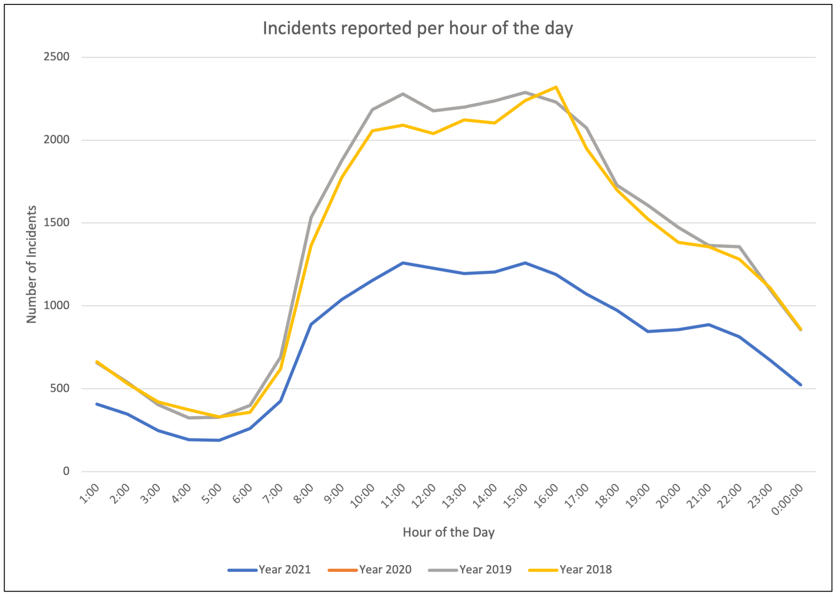

Figure 5: Incident Breakdown by Hour of the Day

3 Methods

3.1 911 Call Priority Model

In order to accomplish our goal of creating the most optimized shift schedule for officers, we need to find the busiest and highest priority days during a week. As described in section two, we used the data provided to us to find the trends on what times of the days and what days of the week were busiest during the last five years.

In addition to determining the busiest days of a week, the priority of the calls that are happening each day is an important factor for determining what days of the week need more officers on shift. However, prioritizing calls can be difficult as many situations in police work are very complicated, and no two 911 calls end in the same result. We came up with a system that we believe gives a very good idea of the priority or importance of a 911 call. We have titled this system the '911 Call Priority Model.'

Section 1.3 describes the data that was provided to us about each 911 call. From this data, we specifically looked at the offenses from the call, the department data of the call, and the progress of the call. We combined these three pieces of information to calculate a numerical score, or "weight" for each 911 call. The higher the score, the higher the priority of that call.

We combined the three data points into a matrix equation to calculate a call's priority score. The setup of this equation is as follows:

$$\begin{bmatrix} C1 \\ C2 \\ C3 \\ ... \end{bmatrix} = a * \begin{bmatrix} OM \end{bmatrix} + b * \begin{bmatrix} DM \end{bmatrix} + c * \begin{bmatrix} PM \end{bmatrix}$$

where C1, C2, C3,... each represent a 911 call. Additionally, OM represents the offense matrix, DM represents the department matrix, and PM represents the progress matrix, all of which will be explained in further detail later in the paper. The variables a, b, and c are the weights of each component of the equation, where each weight determines how much priority is given to the corresponding component of the equation to contribute to the overall priority of the call. After discussing with the SGPD, we have set the weights to be as follows:

$$a = 0.8$$
$$b = 0.1$$
$$c = 0.1$$

The upcoming section will explain the 'Offense Matrix', 'Department Matrix', and 'Progress Matrix' and how we set these up to include them in determining the priority of a 911 call.

3.1.1 Offense Matrix

The Offense Matrix is weighted the highest and thus and most important component in determining a call's priority. In this section we will explain how we created this matrix.

We were provided a list of all of the offense codes from the SGPD. Because there are several offense code for similar cases, similar offense codes were grouped into natures. For example, there are several burglary codes for the varying cases of burglary. Instead of keeping track of each individual code, we put all burglary codes into a single group for burglary. After grouping all offense codes into appropriate groups, we worked with the SGPD to order the groups from most important to least important.

In the offense matrix, each row represents one 911 call, and each column represents a group of codes. The columns are ordered from most important codes to least important codes (left to right). We did this because of the corresponding weights for each column, which will be explained in the next paragraph.

The matrix for offenses is made up of zeroes and ones, with a zero representing that the offense did not happen on that call and a one representing that the offense did happen on that call. If we were to just leave the matrix full of zeros and ones for each offense, they would all equal the same amount numerically. But, there are obviously higher priority codes than others. Thus we multiplied each column with a corresponding weight.

$$\begin{bmatrix} \mathrm{O}_1 & \mathrm{O}_2 & \mathrm{O}_3 & \mathrm{O}_4 & ... & \mathrm{O}_n \\ \mathrm{O}_1 & \mathrm{O}_2 & \mathrm{O}_3 & \mathrm{O}_4 & ... & \mathrm{O}_n \\ \mathrm{O}_1 & \mathrm{O}_2 & \mathrm{O}_3 & \mathrm{O}_4 & ... & \mathrm{O}_n \\ ... & ... & ... & ... & ... & ... \end{bmatrix} * \begin{bmatrix} n/n \\ (n-1)/n \\ (n-2)/n \\ (n-3)/n \\ ... \\ 1/n \end{bmatrix}$$

where n is the number of different offense groups, O_i, $i = 1, 2, 3, ...n$, represents each offense with O_1 being the most important offense and O_n being the least important. The weights range from n/n, or, 1, down to $1/n$, to be multiplied to the corresponding offense row.

There was a total of 60 offense code groups for the SGPD, so the weights ranged from 60/60 to 1/60 from most important code group to least important code group. We used MATLAB to import the 911 call data into a matrix of zeroes and ones and to multiply the weights to the matrix.

3.1.2 Department Matrix

The next portion of the call matrix equation is the Department Matrix. Department data is collected for each call and represents which departments arrived to a particular call – either the law department, the fire department, or the EMS department. We assume that if all three departments arrive, it is a higher priority call than if just one department arrives.

To set up the Department Matrix, we used the same strategy as we did to build the Offense Matrix. Similar to the offense matrix, a one in the LFE matrix represents that the department did show up to the call, and a zero represents that they did not, with each row representing a 911 call. In the department matrix there are only three columns, with one column representing a department.

Since we are focusing on the workload of the law department, we gave more weight to the law department and less weight to the fire and EMS departments. We determined the weights to be 80% on law, 10% on fire, and 10% on EMS. We decided on these weights with the advisement of the SGPD.

The department matrix follows the following format

$$\begin{bmatrix} L & F & E \\ L & F & E \\ ... & ... & ... \end{bmatrix} * \begin{bmatrix} 0.8 \\ 0.1 \\ 0.1 \end{bmatrix}$$

3.1.3 Progress matrix

The Progress Matrix is the final portion of the matrix equation that determines the importance of a call. This data is recorded by the police department for each call and has multiple categories. Below is a table of the different categories of progress and their meanings.

Code	Translation	Explanation
I	In Progress	Time is critical/lights and sirens are commonly used
N	Not In Progress	Time is NOT critical and officers can respond when available
J	Just Occurred	Higher priority that lack a time/urgency factor
Z	Zulu	Placeholder priority that will have a different priority later
A	Alpha	Least urgent medical call priority
B	Bravo	
C	Charlie	
D	Delta	
E	Echo	Most urgent medical call priority
O	Omega	Medical call where no medical resources are required and not sent

Similar to the Offense Matrix, we ordered these categories from most to least important and created a matrix full of either ones or zeros to represent the progress of the call. We worked with the SGPD to weigh each progress category appropriately. The format of the progress matrix is as follows

$$\begin{bmatrix} I & E & D & C & B & A & J & N & Z & O \\ I & E & D & C & B & A & J & N & Z & O \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \end{bmatrix} * \begin{bmatrix} 0.20 \\ 0.20 \\ 0.17 \\ 0.15 \\ 0.10 \\ 0.05 \\ 0.05 \\ 0.04 \\ 0.02 \\ 0.02 \end{bmatrix}$$

3.1.4 Results

After gathering all of the data for each 911 call and putting together each matrix, we were able to combined the matrices for all three components of our equation and get a final numerical score for the overall priority of a call.

$$\begin{bmatrix} C1 \\ C2 \\ C3 \\ \dots \end{bmatrix} = 0.8 * \begin{bmatrix} OM \end{bmatrix} + 0.1 * \begin{bmatrix} DM \end{bmatrix} + 0.1 * \begin{bmatrix} PM \end{bmatrix}$$

In this paper, we will be showing the results for all of the 911 calls that happened in the year of 2017. In order for our results to be helpful in our main goal of optimizing the officer schedule, we calculated the average of the call weight for each day of the week for the whole year. In this model, a day of the week is define to be from 7am on that day to 7am on the next day. For example, calls or accidents that happen between 7am on Monday through 7am on Tuesday count as calls on Monday. We chose to do this with the advice of the SGPD to align the days with the shift schedule for the police officers. Figure 6 shows our results in a bar graph.

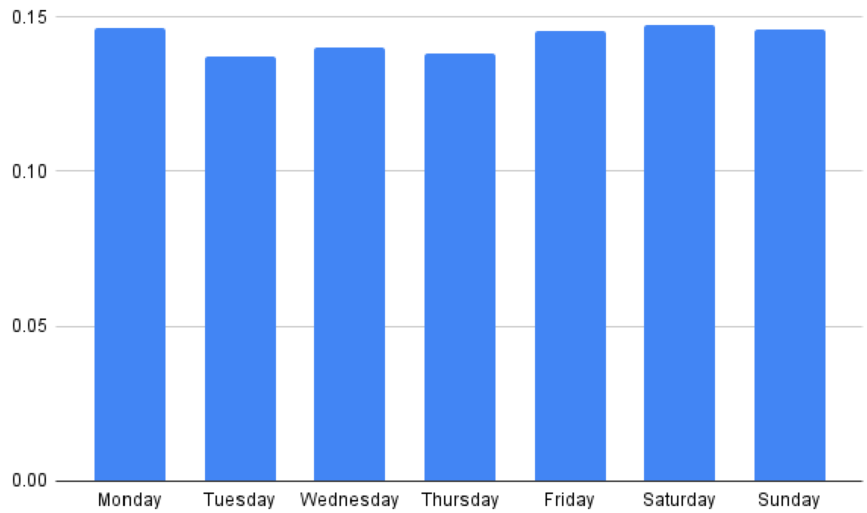

Figure 6: 911 Call Priority Model Results

4 Priority vs Quantity

4.1 Busiest days of the week

Day of Week	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Average Offenses	109.56	109.48	114.12	116.44	117.88	97.23	74.28
Standard deviation	16.62	11.74	12.98	12.11	13.46	13.70	10.60

By calculating the average offenses occurring from Monday to Sunday, We find that Sunday has the lowest average offenses occurred; however, as the call weight we have for those two days, Sunday has almost the same Weight-Score as other days.

In order to find the reason of this, we analyzed the offenses happened on Friday and Sunday. There are more severe offenses(for example, family offense,Burglary/Breaking Entering and so on) occurring on Sunday.

4.2 Distribution of offenses

We also analyzed the trend of offenses happening from Monday to Sunday. We made box-plots for each day of week. By looking at the distribution of offenses,the medians of each day are represented by the line across the middle of the boxes, and the whiskers are the minimum and maximum numbers of offenses.

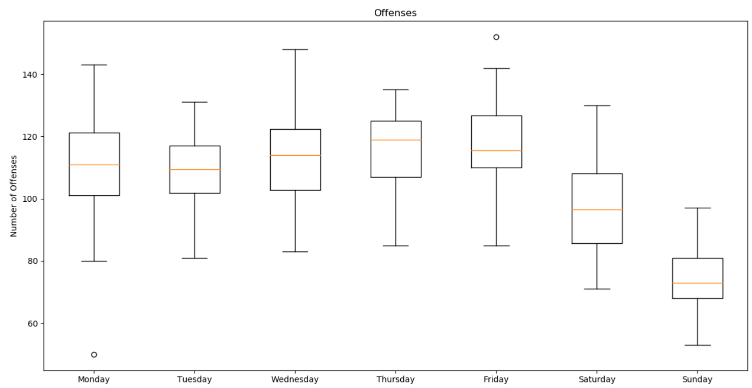

Figure 7: Distribution of Offenses Box Plot

We assumed the offenses followed normal distributions. In order to verify this assumption, we used Shapiro-Wilk test. The reason we choose this method is because Shapiro-wilk test is indicated as the best test to determined whether the date comes from a normal distribution or not. The results showed that at significant level $\alpha = 0.05$, there is not enough evidence to say the distribution of offenses on six of the days are not normally distributed (with Thursday as the exception). Therefore, we could use the normal distribution formula to calculate the probability of the offenses:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

where μ represents the mean of offenses and σ is the standard deviation of the offenses.

5 Day Priority Model

As discussed in the section 3, after creating the call matrix to get a weight for each individual call, we then separated the data into days of the week to analyze which days are typically higher priority (see Figure 6). When we switched from an individual 911 call format to a day of the week format, we were able to include accident and citation data, which accounts for the officers time addressing traffic issues. This data is important to include in the overall workload for each day of the week.

The accident and citation information was provided to us from the SGPD because they thought it would be a great addition to our previous results to make them more accurate. The accident and citation information is the information of the traffic accidents that the SGPD responded to and the citations that they gave out when patrolling the road and pulling people over. This is separate from the 911 call data because there is typically not a 911 call for traffic violations that police respond to.

To include the accident and citation information, we simply found the number of accidents/citations that occurred for each day of the week for the entire year of 2017. Then, we averaged that number and found the proportion of the average over the total number of accidents/citations.

We added the results for both the accidents and citations to the average 911 Call Priority Model results that we presented in the previous section.

$$\begin{bmatrix} Monday \\ Tuesday \\ ... \\ Sunday \end{bmatrix} = 0.8 * \begin{bmatrix} 911CP \end{bmatrix} + 0.1 * \begin{bmatrix} ACC \end{bmatrix} + 0.1 * \begin{bmatrix} CIT \end{bmatrix}$$

We decided (again, with the help of the SGPD) to weigh the 911 Call Priority Model as 80% of the total priority of a weekday, and the accident and citation data each weighed as 10% each. Our results for the priority of each day of the week for 2017 are shown in Figure 7.

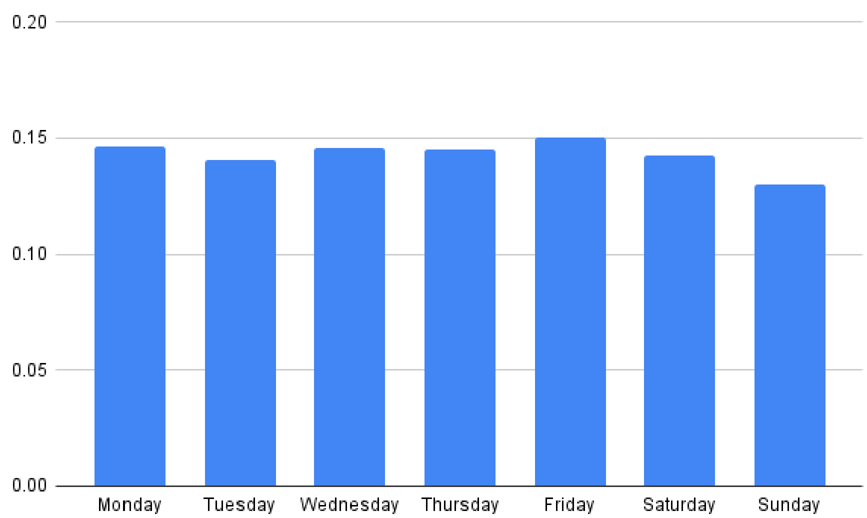

Figure 8: Day Priority Model Results

Keep in mind that the days in this model are from 7am to 7am the following day due to the way the shifts are for the police officers. The chart shows that when considering five data points (offense, departments, progress, accidents, and citations) in the workload of a day, **Friday has the highest overall workload, while Sunday has the lowest overall workload**. We have not yet looked into how much of a difference of workload is represented by a small number, such as 0.001. We have the question of "how much does a 0.001 difference in score relate to the difference in workload?". It is possible that a very small difference, such as 0.001, might actually represent a huge difference in workload. This is something that we plan to look at in the future to better measure the differences of workload from day to day.

6 Conclusion

From unorganized data sets, to data driven results, our project has been able to provide useful information that can be easily transitioned into department decision making. To reiterate our goal at the start of our project, we wanted to help the police department utilize their current resources in the most efficient way possible so that they could allocate resources in a better way that would help keep the community safe and give their employees proper help during more strenuous times. In our work, we used matrix algorithms to give a numerical score for each day which has provided us a way to determine which days require the most resources, as well as days that may not require as many resources and thus can help relieve the employees of being overworked. As a result we have been able to find optimal days to have overlapped schedules to help with busier days. As a result we aspire that our results will help the police department utilize their resources efficiently while maintaining and increasing the safety of the community.

7 References

Anderson-Darling and Shapiro-Wilk tests - NIST. (n.d.). Retrieved April 11, 2022, from https://www.itl.nist.gov/div898/handbook/prc/section2/prc213.htm

Curtin, Kevin M., et al. "Determining Optimal Police Patrol Areas with Maximal Covering and Backup Location Models - Networks and Spatial Economics." SpringerLink, Springer US, 17 Oct. 2007, https://link.springer.com/article/10.1007/s11067-007-9035-6.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, 10(2), 486–489. https://doi.org/10.5812/ijem.3505