Optimize Physician Staffing in the Emergency Department

David Izrailov, Katherine Liu, Raul Mizan, Nika Nasseri, Haider Zaidi
Schulich School of Business, York University
OMIS4000: Prescriptive Analytics
Professor Raha Imanirad
April 4th, 2023

Introduction:

There is an expected influx of patients over the span of three days (Monday-Wednesday) and the hospital's emergency department needs to assign physicians for the three shifts of each day. The objective is to minimize the length of stay (to measure physician speed) and minimize the rate of return (to measure physician quality) for each patient. The problem of solving this question of who to staff is solved using a binary optimization approach with the only decision variable being which physicians to staff and when. The approach for approximating demand was conducted using statistical approximations of previous demand data to provide a continuous demand function for each shift-day pair. This creates a continuous range for estimation and allows for management to create staffing schedules based on influx severity. The problem states a variety of constraints such as making sure no physician is staffed twice in a row. However, upon further analysis of the dataset, more constraints are formulated to keep in mind physician preferences and well-being, alongside demand constraints as well. While some may choose to prioritize speed or quality, it was decided that both are equally relevant and neither should be prioritized. In the event management decides to focus either speed or quality, coefficients may be added to the model to adjust weighted to the desired amount.

Data Collection/Analysis and Model:

Data Relevancy:

Upon inspection of the dataset, features included Date, Physician ID, Average Patient Age, Female Patients (%), Average Patient ESI, Average Patient LOS (minutes), Average Lab Order Count, Average Radiology Order Count, Average CT Order Count, 72-hr Rate of Return, Shift, Day, Month, Year, ED Volume, Physician Speed (LOS in minutes), and Physician Quality (72-hour return rate). Features were assigned to three buckets:

1) Irrelevant: Date, Average Patient Age, Female Patients (%)
2) Supplementary: Average Patient ESI, Average Patient LOS (minutes), Average Lab Order Count, Average Radiology Order Count, Average CT Order Count, 72-hr Rate of Return
3) Crucial: Physician ID, Shift, Day, Month, Year, ED Volume, Physician Speed (LOS in minutes), and Physician Quality

The basis behind what was deemed crucial were features that are directly related to the model. Supplementary helped derive additional insights, and irrelevant were unused in either the model or insights.

Prioritization:

While the intent was to assign physicians based on an influx, it was decided that both quality and speed of physicians are equally important. Financial impacts were given the lowest priority as human lives are at stake. It is important to not only prioritize speed, but to provide the best care for patients possible. As such, equal weighting was given to both. Understably, without full prioritization of speed, it is expected that more physicians will need to be staffed than if the model was based purely on efficiency. While this is the case with many optimization models, these are not machines producing goods, these are physicians saving lives. Upon analysis, the tradeoff between speed and quality was not heavily influenced for the majority of physicians, however some did display a clear tradeoff. Physicians 18 and 19 show to have a lower ROR with respectively higher LOS. This dynamic can be seen in Exhibit 1.1 and 1.2.

Objective Function:

The objective function intended to minimize both the time spent and rate of return for each patient. This was done using only one binary decision variable representing a given physician being staffed on a certain day and shift. The objective function can be seen below. It incorporates 4 major elements. Variable Q_i represents the quality (ROR) associated with each physician i, and variable S_i likewise represents the speed (LOS) associated with each physician i. Decision variable $D_{i,j,k}$ represents a binary decision to staff a physician i on shift j of day k. Finally, the $E[X(j, k)]$ term represents the expected value of demand for a given shift and day pair which is determined using statistical approximations of demand. The objective function minimizes the length of stay and rate of return for each patient by assigning various doctors subject to a series of constraints to be mentioned.

$$min\ z = \sum_{i=1}^{28}\sum_{j=1}^{3}\sum_{k=1}^{3}(Q_iD_{i,j,k} + S_iD_{i,j,k}) \times E[X(j, k)]$$

Approximating Demand:

A key element of the model is this notion of expected value in a given shift-day pair. The first step in acquiring such values was aggregating the data in such a way that the total ED Volume for each individual shift was calculated and compared to other instances of itself. For example, all the shift 1, day 1 events had their ED Volume summed and appended into a list of 130 elements. Utilizing this list with a single dimension, a histogram can be created to get the distribution of ED Volume for a shift-day pair. A statistical distribution can be then fitted to these histograms and an approximation of demand for each shift-day pair can be obtained. Reference Exhibit 1.3 for the various statistical distributions fitted to each scenario.

Tunable Parameter:

These fitted distributions represent the probabilities of ED Volume in a certain scenario. Utilizing this property, a tunable parameter was created named “Percentile”. This parameter paired along with these distributions can approximate the expected ED Volume based on what percentile event is expected. Information on the severity of the influx the ED expects was not given. However with this parameter, if they expect a 90th percentile event influx (the next observation will be greater than 90% of all other observations), they can set this parameter to 0.9 and the expected ED Volume will be generated. For the purposes of the model, a 90% event was assumed; however it is encouraged that management to change this parameter to fit their expert opinion. For their reference, a sensitivity analysis was conducted to show the difference in expected demand based on the percentile changes as shown in Exhibit 1.4. This parameter alongside the fitted distributions gives us the $E[X(j, k)]$ portion of the objective function.

Physician Speed and Quality Statistics:

The last portion of the objective function requires the individual speed and quality values associated with each physician. This was calculated by taking an average of the Physician Quality and Physician Speed columns in the dataset provided. Exhibit 1.1 and 1.2 shows each of these values as well as a breakdown by shift. It is clear that speed and quality are stable between shifts meaning a physician will have the same performance regardless of if staffed in the morning, afternoon, or night. These values were taken and added to lists to be used for the objective function.

Additional Considerations:
One addition that can be made to the objective function is the introduction of coefficients for speed or quality. These affect the weighting of each to influence the decision in a certain way. It was decided to maintain the tradeoff between the two at a 1:1 ratio without additional information from management and thus no coefficients were introduced to bias the model.

Alongside these considerations, there were other options for models available such as stochastic and goal programming. Stochastic would use the distributions to evaluate all possible scenarios; however it is believed that management would be able to estimate influxes reliably well and thus would have no need for such a complex model. They can rather input influx severity as a parameter and produce scheduling as such. Stochastic models are complex and hard to interpret for those without a background in optimization and statistics. Goal programming was also an option, which provided the ability to add soft constraints. The goal of the ED is simply put as staff optimally keeping in mind speed and quality. The binary model addresses this goal well with minimal complexity. While soft constraints may have been helpful for fairness constraints, the model still performs well with all constraints being hard.

Constraints:

Constraints have been divided into four main groups: Employment, Demand, Physician Specific, and Labour/Fairness.

Employment Constraints:
These constraints encompass the staff that no longer work at the ED and should not be staffed under any circumstance. Exhibit 1.5 shows the physicians who have worked in the past 6 months. Physician 3 and 20 are not in the dataset meaning they have not worked in the past two years and are assumed to no longer be employed. Physicians 22 and 28 were also found to have not worked in the past six months. It is assumed that they no longer are employed in this ED. The constraints are below.

$$x_{3,j,k} = 0, \forall j \, \forall k, \qquad x_{20,j,k} = 0, \forall j \, \forall k, \qquad x_{22,j,k} = 0, \forall j \, \forall k, \qquad x_{28,j,k} = 0, \forall j \, \forall k$$

Demand Constraints:
Given the nature of the model, since all physician quality and speed statistics are positive numbers and the model aims to minimize, it will attempt to staff as few physicians as possible. To counteract this, constraints were added to state that the sum of all physician capacity should at least exceed expected demand $E[X(j, k)]$ in each scenario. The capacity value for each physician was obtained by calculating the average ED volume per shift per physician, as shown in Exhibit 1.6, and taking the maximum value of the three shifts. In every case, it was the shift 1 value. The shift 1 average ED per physician was appended into a list to be used as the value each physician is expected to be capable of handling. It was assumed that if they could handle an amount in the morning, they could do the same on any shift. Additionally, the average ED volume was taken over the upperbound ED volume to represent capacity so as to not overwork each physician, ensuring quality is not sacrificed, and acts as a buffer in the case that there is more demand. The constraint was formulated to state that the sum of all chosen physicians and their given capacity are greater than expected demand. The mathematical formulation can be seen below.

$$\sum_{i=1}^{28} (x_{i,j,k} * C_i) \geq E[X(j, k)] \, \forall j \, \forall k$$

Physician Specific Constraints:
These sets of constraints refer to certain preferences which have been found through analyzing Exhibit 1.7 and 1.8. When analyzing the counts of shifts for physicians on the days of interest, it becomes evident that some physicians have certain shifts they have worked very few of in the past two years. This is likely due to an availability/preference and constraints have been set in order to be faithful to these preferences. There are a select few physicians that choose not to work the second or third shift, however all must be available for shift 1. In terms of shift availability, the constraints can be seen below.

$$x_{7,2,k} = 0\ \forall k, \quad x_{7,3,k} = 0\ \forall k, \quad x_{13,3,k} = 0\ \forall k, \quad x_{16,2,k} = 0\ \forall k,$$
$$x_{16,3,k} = 0\ \forall k, \quad x_{18,2,k} = 0\ \forall k, \quad x_{18,3,k} = 0\ \forall k, \quad x_{19,2,k} = 0\ \forall k,$$
$$x_{19,3,k} = 0\ \forall k, \quad x_{21,3,k} = 0\ \forall k$$

The same constraints occur for specific days of the week with the dynamics shown below. Note that k=1 for monday, 2 for tuesday, and 3 for wednesday.

$$x_{1,j,1} = 0\ \forall j, \quad x_{1,j,2} = 0\ \forall j, \quad x_{7,j,1} = 0\ \forall j$$

Labour and Fairness Constraints:
This set of constraints are in-place to make sure the model does not constantly staff the same set of physicians and that physicians are not overworked. The ED department has stated that they would like no physician to be staffed twice in a row. Constraints were added to say each physician may only work once a day, as well as a constraint that no physician should be staffed on shift 3 and shift one the following day. The combination of these two constraints makes it so no physician can work twice in a row while also making sure they cannot work twice a day as well. It is understandable that optimization is a priority however these are people and not machines. Consideration into physician well being was a top priority in the model. The two constraints are seen below.

$$\sum_{j}^{3}(x_{i,j,k}) \leq 1, \forall i\ \forall k\ , \qquad x_{i,3,k} + x_{i,1,k+1} \leq 1, \forall i,\ k = \{1, 2\}$$

Our analysis has also concluded there are two separate classes of physicians: regulars and reserves. The regulars come in much more frequently than reserves and it was assumed that the response rate from regular physicians would therefore be higher. As it is a quick turnaround time to staff and let the physicians know when they are working, it is best to reach out to the reserve physicians only when necessary and not over rely on their availability. Therefore the constraint was added that the reserve doctors may only be selected at most once for this shift. A graphic showing the counts of each physician's shifts can be seen in Exhibit 1.7. The reserve physician constraint(s) can be seen below

$$\sum_{j}^{3}\sum_{k}^{3}(x_{i,j,k}) \leq 1,\ i = \{7, 13, 14, 16, 18, 19\}$$

Analysis also showed clear patterns for days physicians worked. As these are the livelihoods of many physicians, if there is such an instance of a clear pattern to work on a day it is granted. The only instance of this was Physician 1 working almost every Wednesday for the past two years as shown in Exhibit 1.8. A constraint was added to make sure he is staffed on Wednesday.

$$\sum_{j}^{3}(x_{1,j,3}) \geq 1$$

Other Considerations:
Most of the ground is covered in terms of necessary constraints. The model has addressed the demand requirements of the ED as well as the preferences of physicians, all while prioritizing the well-being of the physicians and paying close attention to work-life balance. Some considerations for the future would be to spend more time looking at the livelihood component of the physicians. The focus would be on understanding the previous staffing patterns which physicians are accustomed to and simply adding physicians if ED volume is expected to surge. It is understood that the model may disrupt the scheduling pattern many physicians have become accustomed to and with more time would focus on centering the model around these accustomed schedules. It was also considered adding a constraint to link the last shift in the dataset and the scheduling of Monday-Wednesday. Upon further analysis the last point in the dataset is July 1st 2016 which is a Friday meaning this is not truly the latest datapoint.

Additional Analysis

ESI and Financial Insights:
To gain a deeper understanding for the operating environment of the ED, analyses were conducted to assess the relationship of different variables and each physician, to determine if outliers are present (or lack thereof), which can be further investigated, alluding towards inefficiencies. As shown in Exhibit 2.1, it was concluded that the average patient ESI among all physicians**,** were almost the same, at approximately 2.93, indicating no one physician undertook more or less urgent cases. Discrepancies began to arise in the average order counts as shown in Exhibit 2.2, 2.3, 2.4: Physician 17 had considerably higher radiology and CT order counts than other physicians. The relationship between the Physician LOS and the Patient LOS was also conducted in Exhibit 2.5, where the delta of the average Physician LOS and the patient's perspective average Physician's LOS to assess the significance of factors impacting a Physician's LOS versus their patient's; likewise, with the 72-hour ROR. Concerning LOS, Physician 24 had a considerably higher LOS delta, in some cases, twice the size or more, compared to other physicians. Physician 24's LOS was much higher than their patient's. Further investigation must be conducted to assess the cause.

Speed and Quality Tradeoff:
For the majority of physicians, there is no direct trade-off between the speed (LOS) and quality (ROR), however, for a few, a positively correlated trade-off is present. Physicians 18 and 19 each have a single shift, with the first and second highest LOS' and the lowest ROR (i.e., highest quality) respectively. A similar but marginally weaker relationship exists with Physician 16, however, this Physician has two shifts instead of one; most have three. It is evident that Physicians with fewer shifts, and a longer LOS, have a lower 72-hour ROR, which may be a result of them being spread less thin across shifts and paying closer attention to patients in their care.

Recommendations

The model was created to be as robust as possible to allow for adjustments from the department's managerial team to make the final decisions. The department is recommended to prioritize defining an "influx" of patients and choosing a constraint set to finalize staffing schedule.

Management should then look into the lab, radiology, and CT order inconsistencies and see if additional actions/training is needed.

Finalize Staffing Schedule
As previously mentioned, a 90th percentile event was assumed to occur, however this assumption should change based on management's or an expert's observations on the event. After this change, management should choose which of the three constraint sets is most appropriate.

As shown in Exhibit 3.1, three constraint sets were created that ranged from the most optimized set to the most fair set. Taking into account the advantages and disadvantages of each set, it is recommended that the department work with set 2 as it could increase responsiveness from the physicians during this quick turnaround period. By considering physicians' previous history on how often, when they were, and when they were not staffed, there is a higher likelihood that these physicians will be available to be staffed in this schedule. Therefore, constraint set 2 that can increase responsiveness while still prioritizing speed and quality is the recommended set. The department should also staff Physician 1 to a Wednesday shift as while it is not necessarily a preference constraint, Physician 1 has worked almost every Wednesday for the last 2.5 years and there is a very high likelihood that they would be available for that shift this week.

Based on an assumption of a 90th percentile event and using mainly constraint set 2 (added fairness constraint for physician 1 working Wednesdays), a recommended staffing schedule was created as shown in Exhibit 3.2. Immediate action should be taken to notify physicians of the schedule and adjustments be made for any changes to physician availability.

Business Insights
When analyzing the rest of the dataset beyond the core model, there is a concern towards the inconsistency of additional tests needed such as the lab, radiology, and CT order counts. In particular, Physician 17 has a marginally higher count across all three types of orders and management should have further discussion as to why that may be, especially since they rank one of the highest in combined speed and quality. Management may consider additional training to ensure orders are being made appropriately.

Relevant Limitations and Implications
There are limitations and accompanying implications associated with the model such as hindsight bias, the weighting of quality and speed, ethicality of essentially ranking physicians, the inability to account for social synergies affecting speed and quality, and the inability to factor in resource and financial constraints. In terms of hindsight bias, it is very easy to look back after the three days and be able to adjust the percentile parameter to the actual percentile event observed and add constraints to better represent physician availability. Thus, being able to produce a more optimal staffing schedule post-event than predicting and producing a schedule pre-event. This hindsight bias is a limitation to most models as very rarely will a model have perfect information.

As mentioned in the objective function, the model assumes an equal weighting between the speed and care for patients and there is a limited amount of information to determine if the hospital would like to prioritize one or the other. The weighting is important to determine as prioritizing speed typically would result in a higher risk of lower quality while a larger focus on quality could result in longer LOS and increased risk of resource constraints and additional patients kept waiting. The weights can easily be modified using coefficients in front of the variables.

The model itself will also yield information to rank the physicians by speed and quality in order to determine staffing. Based on the staffing schedule in Exhibit 3.2, 4 physicians who are consistently staffed could be considered the “top performers” while the 8 physicians not staffed could be “underperformers” based solely on speed and quality. Speed and quality are only two indicators of a wider range of performance metrics and while performance evaluations are ethical and expected, to rank physicians against each other to determine an specific outcome could be considered unethical and something that the physicians themselves would not like or consent to.

Another implication includes the inability to account for social synergies affecting the speed and quality of physician work. Synergies could be in place where specific sets of physicians working together could motivate and increase quality and speed. However, the alternative can also exist where specific physicians could have conflicts or decrease quality and speed. As the available data is primarily quantitative, it would be difficult to factor in the more qualitative considerations unless regression models were run for each physician and physician they have previously worked with to draw correlations.

Lastly, there is no data present to add resource and financial constraints such as room capacity, budget, and equipment. To fully be able to meet an influx of trauma patients, it is important to know the current number of patients in the emergency department, the physical capacity of the hospital, the available and expected demand for equipment, and if there are any budget constraints that could limit the quality and speed of care. Additionally, it is unknown if there is time allotted to sanitizing and preparing each room for patients that is an additional time constraint.

Conclusion

The model created is easily deployable, scalable, adjustable, and agile. It provides quick results using statistical distribution which gives it the element of robustness. It was also created in a way that allows for easy managerial modification if need be and is applicable to scheduling outside of the 3 day period if modifications are made. The tunable parameter paired with fitted distributions allows for a precise estimate of demand for a given shift-day pair. The model was designed to keep in mind physician well-being as well as not sacrificing speed for quality and vice versa. A staffing schedule was produced based on a 90th percentile influx event and the aforementioned constraints but it is recommended that management update accordingly to the level of influx expected as well as which set of constraints to apply to the model.

Additional insights were discovered such as physician inconsistency in lab, radiology, and CT counts as well as seeing certain physicians which exhibit a speed-quality tradeoff. The recommendations going forward are to further investigate the order inconsistencies and those who exhibit the speed-quality tradeoff.

Appendix

Exhibit 1.1: Average Physician Speed per Shift

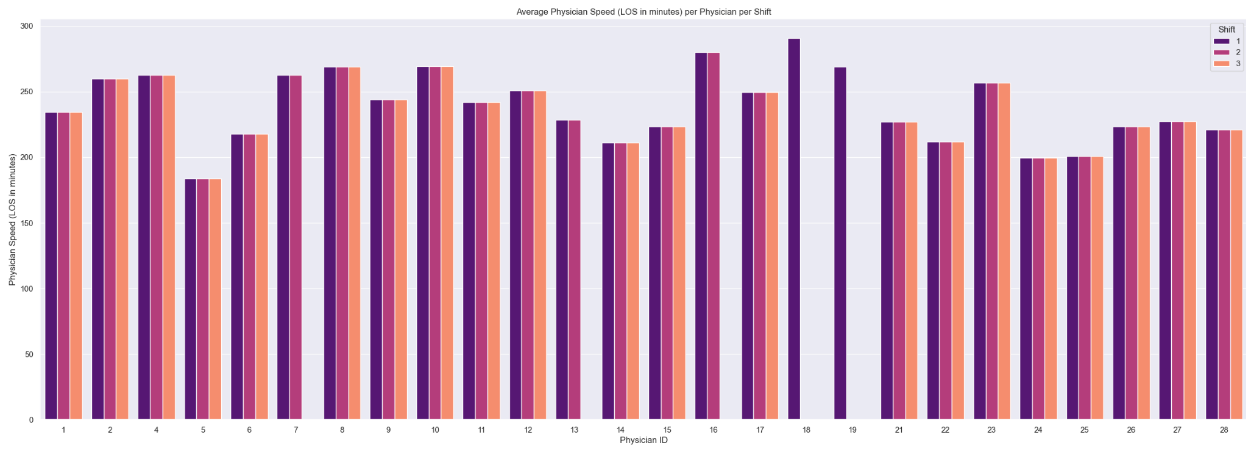

Exhibit 1.2: Average Physician Quality per Shift

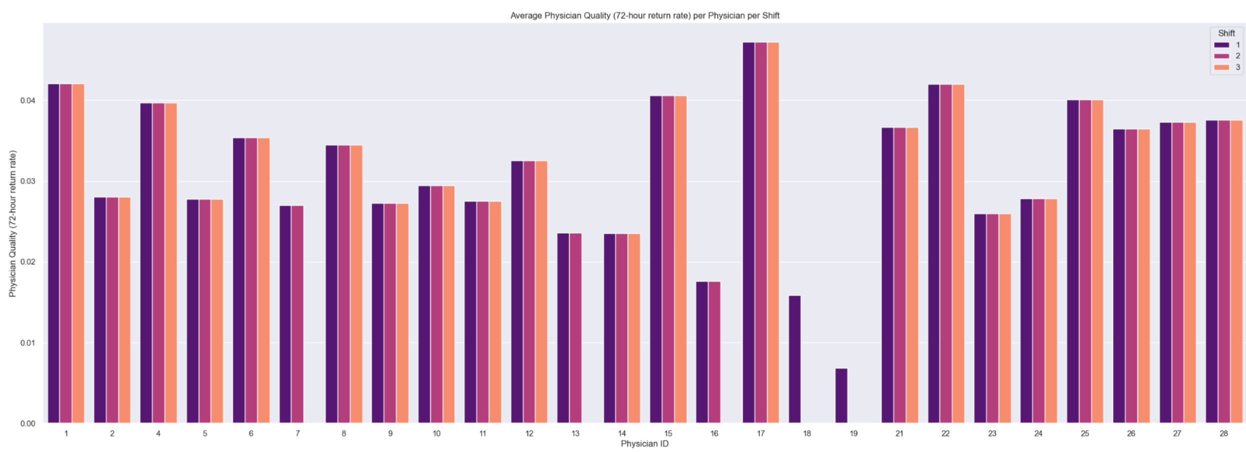

Exhibit 1.3: Distributions Used for Each Shift-Day Pair

Distributions used for each shift-day pair			
	Day 1	**Day 2**	**Day 3**
Shift 1	Gamma	Gamma	Gamma
Shift 2	Gamma	Gamma	Logistic
Shift 3	Lognormal	Lognormal	Lognormal

Exhibit 1.4: Sensitivity Analysis

Sensitivity Analysis of expected demand					
	English Definition	45th percentile	60th percentile	75th percentile	90th percentile
$E[X(1,1)]$	Monday Morning	1326	1613	1982	2615
$E[X(1,2)]$	Tuesday Morning	1174	1418	1723	2224
$E[X(1,3)]$	Wednesday Morning	994	1236	1555	2117
$E[X(2,1)]$	Monday Afternoon	472	574	706	934
$E[X(2,2)]$	Tuesday Afternoon	416	504	619	818
$E[X(2,3)]$	Wednesday Afternoon	446	527	619	764
$E[X(3,1)]$	Monday Night	46	75	132	294
$E[X(3,2)]$	Tuesday Night	47	79	139	317
$E[X(3,3)]$	Wednesday Night	45	77	142	339

Exhibit 1.5: Shift Counts Over the Last Six Months

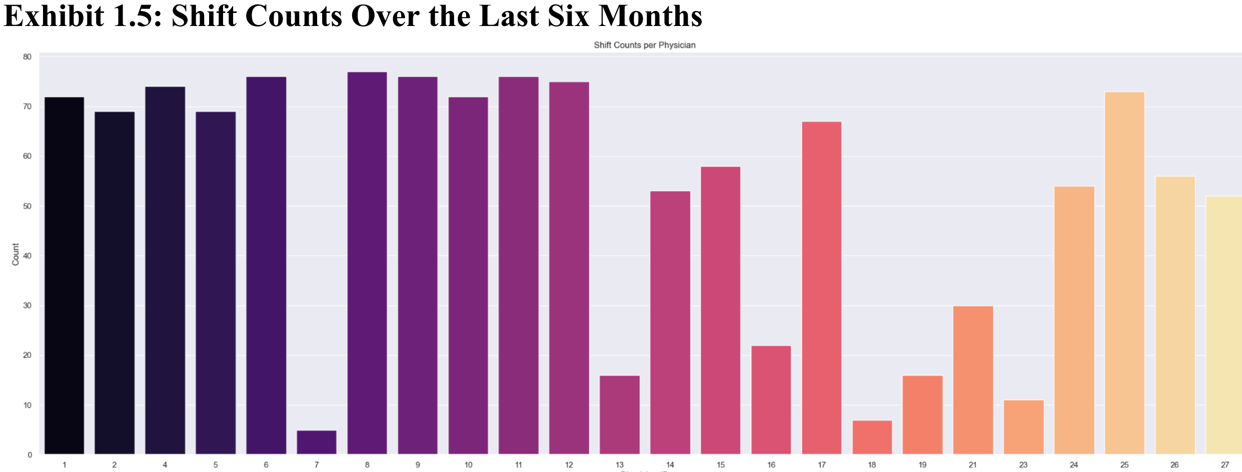

Exhibit 1.6: Average Physician Capacity (ED Volume) per Shift

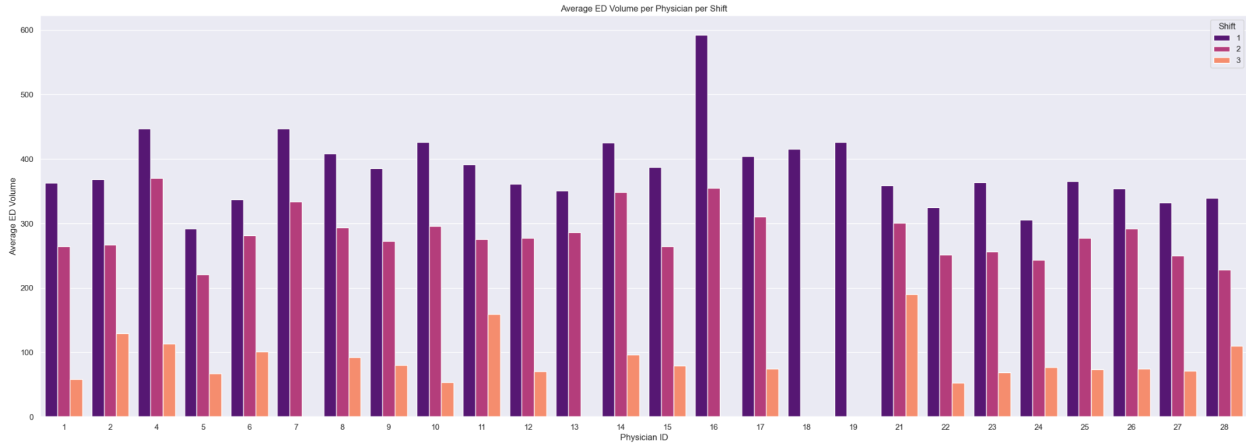

Exhibit 1.7: Counts of Each Shift per Physician

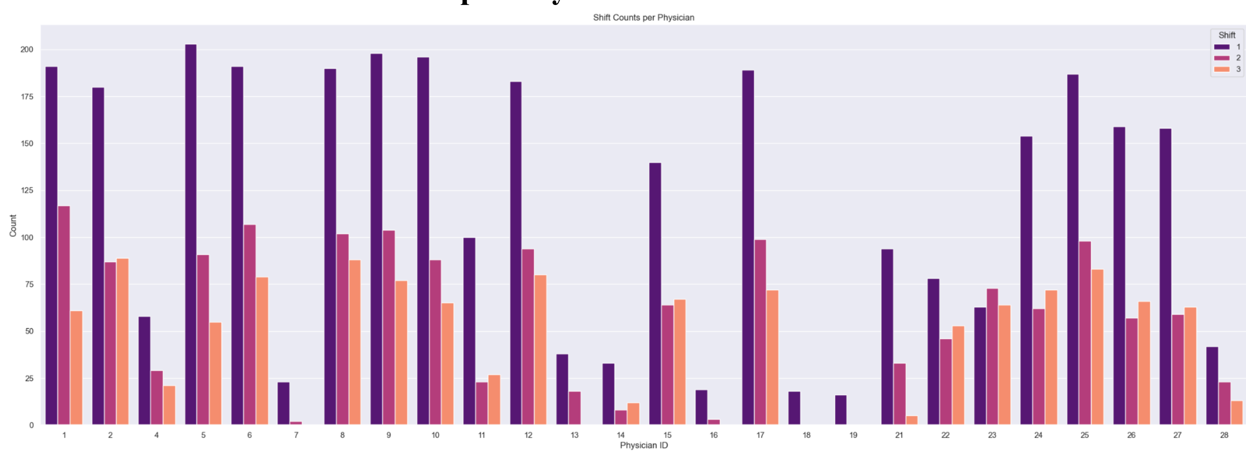

Exhibit 1.8: Counts of Each Day per Physician

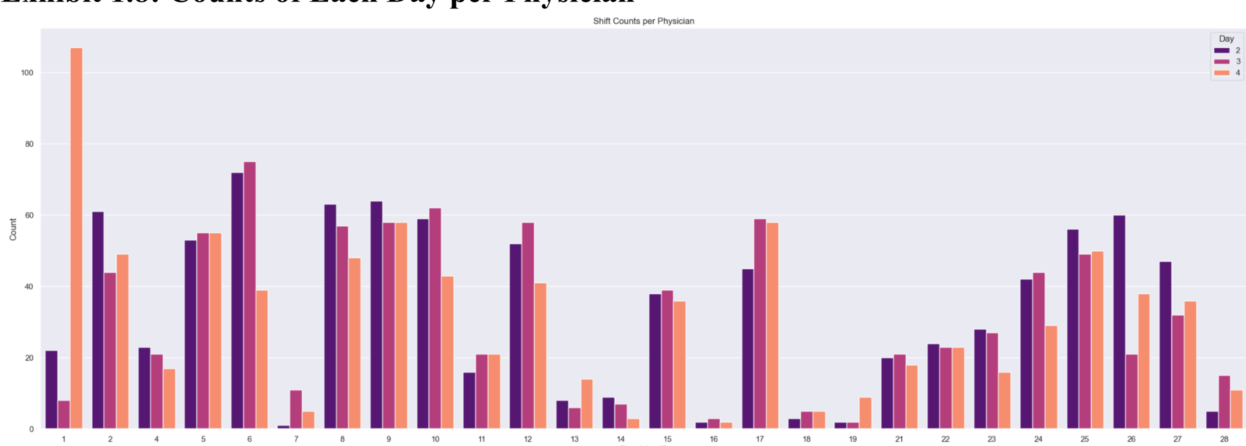

Exhibit 2.1: Average Patient ESI per Physician

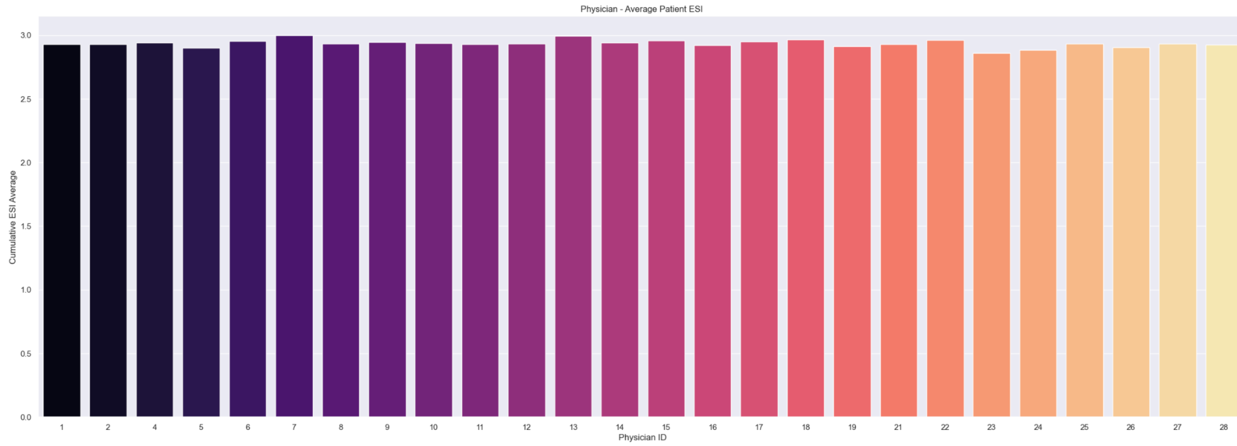

Exhibit 2.2: Average Lab Order Count per Physician

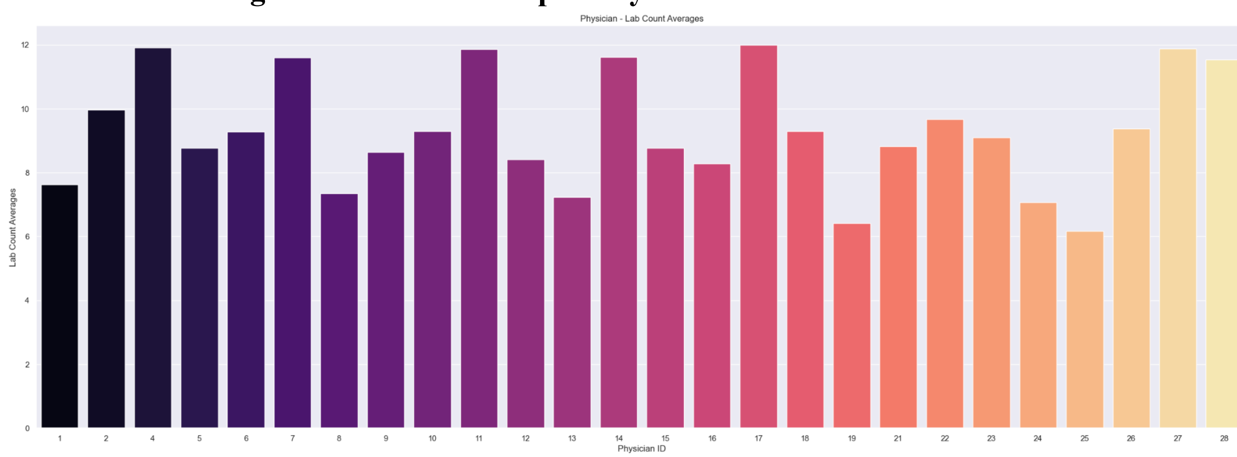

Exhibit 2.3: Average Radiology Orders per Physician

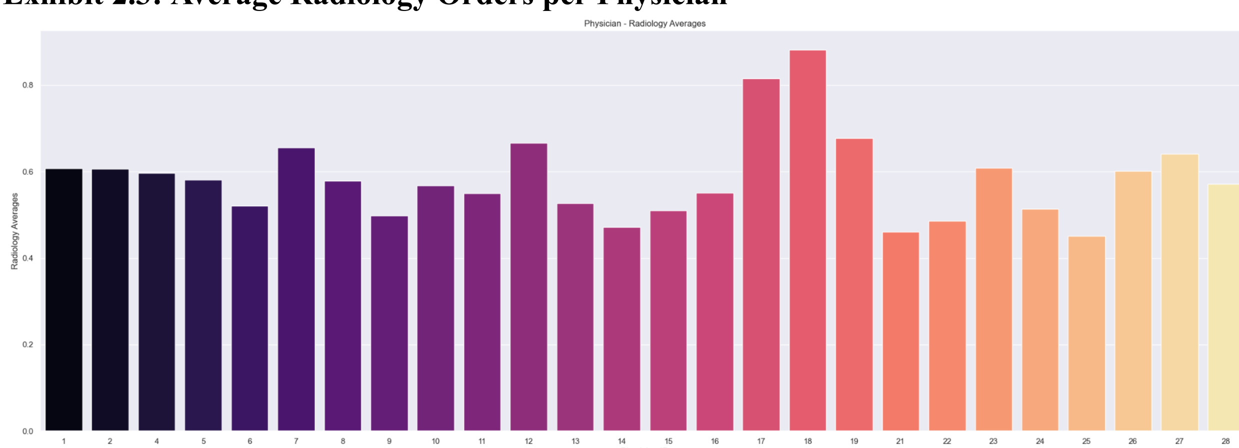

Exhibit 2.4: Average CT Orders per Physician

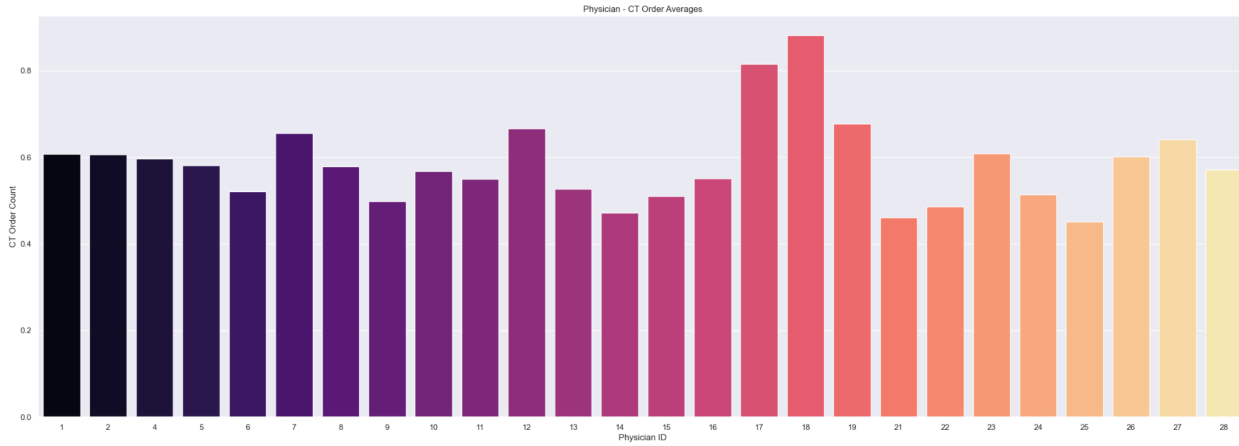

Exhibit 2.5: Physician vs Patient LOS

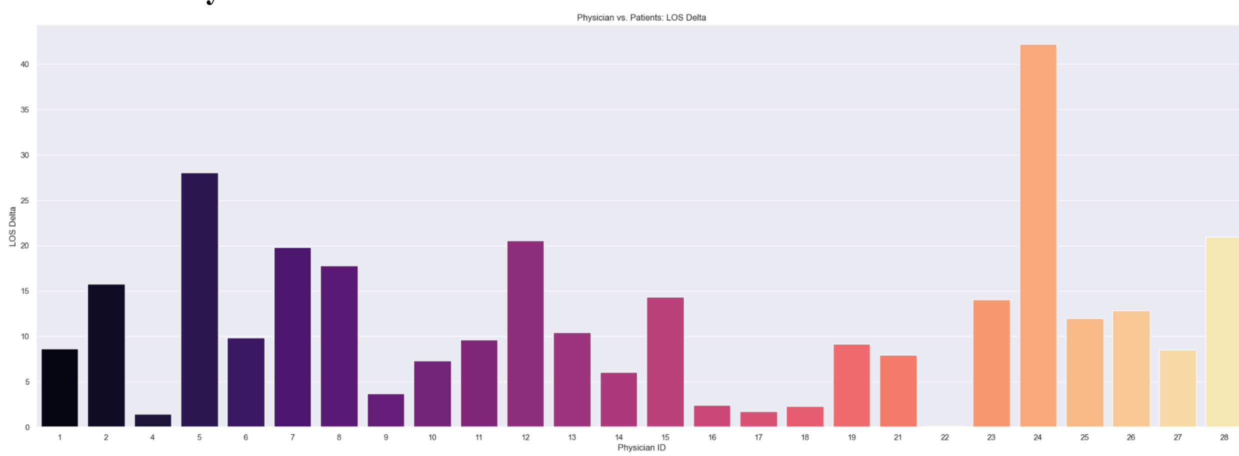

Exhibit 2.6: Physician vs Patient ROR

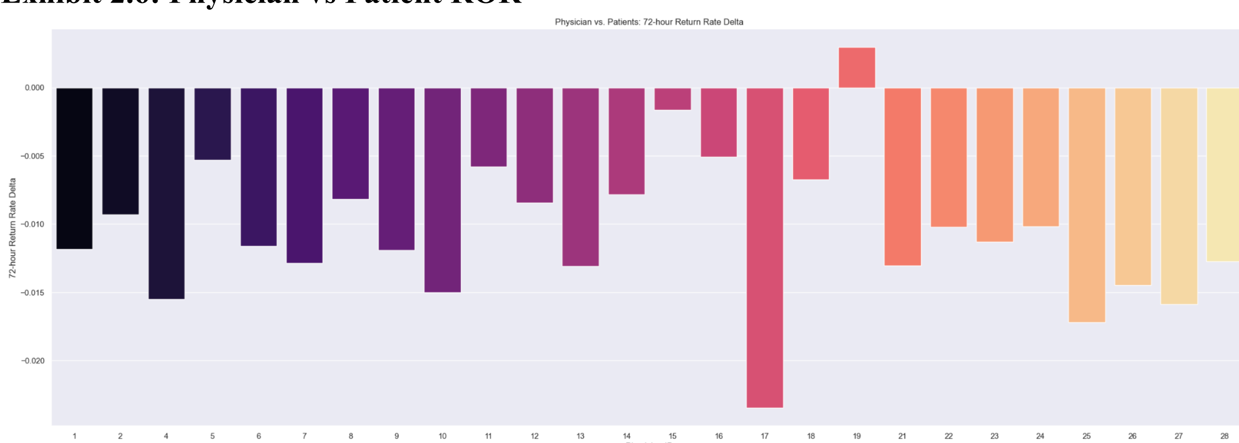

Exhibit 3.1: Constraint Sets

Constraint Set	Constraints Included	Advantages	Considerations
1 (Optimized)	Minimal constraints that include which physicians are no longer with the department, demand constraints, and consecutive shift constraints.	This set prioritizes quality and speed the most to deliver the best service with minimal additional constraints that would diminish those factors.	Is rather inhumane as the model will consistently pick the top physicians on rotation without considering any other factors.
2 (Preference)	On top of Set 1, includes physician preferences/availability in terms of day and shift as well as limiting reserve physicians to one shift during the period.	Still prioritizes speed and quality and reduces the chances of physicians being unavailable as it takes preferences into account. Reserve physicians would still be reached out to but with limited expectation.	A number of physicians would not be staffed during this period while other physicians are staffed multiple times which could be seen as unfair.
3 (Fairness)	On top of Set 1 and 2, includes fairness constraints for preferred day/shift, one shift per physician, and no reserve physicians.	Ensures that no one physician is "overworked" and shifts are as evenly distributed as possible taking into account the quality and speed. Additionally, reserve physicians would not need to be reached out to.	As it is an influx scenario with a quick response needed that prioritizes quality and fast physicians, it may not be necessary to consider all these additional constraints.

Exhibit 3.2: Schedule of Physicians

	Day 1			Day 2			Day 3			
Physician/ Shift	1	2	3	1	2	3	1	2	3	Count
1										1
2										0
4										3
5										2
6										2
7										1
8										0
9										0
10										2
11										1
12										0
13										0
14										1
15										3
16										1
17										3
18										0
19										1
21										0
23										0
24										1
25										3
26										1
27										1

Link to HTML files and Code: [Here]