Analysis

Time and costs are often highly correlated for operations carried out by organizations, and an effort to identify inefficiencies in operational processes can help make business decisions with respect to optimization, improvement and resource allocation.

In this report, I analyze the given data with the aim to support decision making for optimizing the time taken for patient treatment. I do so by using statistical insight based on Gaussian distributions, and correlation models for treatment time.

The essential idea in using these tools is as follows:

If the distribution of a time variable has a Gaussian distribution with a narrow bell cure (i.e. low standard deviation), the process is better controlled because there is specified narrow time domain within which most patients will be treated. Conversely, if their distribution has a high Std. Deviation and variance, the process is not well controlled: time taken by the process varies greatly implying the process is not well defined and has inefficiencies. To identify, the causes of inefficiencies in the process, time distributions by categories are plotted to infer whether the category has a uncontrolled sub-process which is adding to the overall inefficiency.

In essence, this report aims to answer the following questions from the data:

1. Is the process of treating patients controlled with respect to the time it takes?
2. Is the process efficient, or inefficient with respect to the time it takes?
3. Are there specific categories for patients/complaints for whom the process is more inefficient?
4. How did the inefficiency impact the status of patients' treatment, giving a measure of urgency in pursuing the process improvement?

Summary Analysis

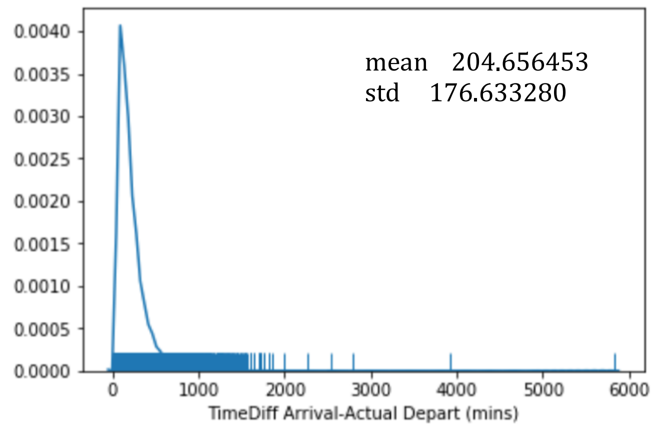
Before data was analyzed, a quick summary analysis was performed to gain some insight on the raw data given. To do this a high-level summary on the data was performed and has been included in Appendix A.

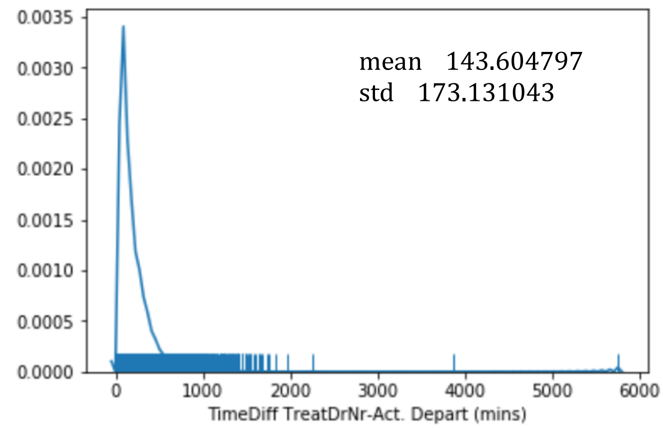
*See **Appendix A** for Summary Analysis.*

Analyses

Interval Distribution

The data had two variables which indicated time related information about the process. They were distributed as follows:





The figures indicate that there is some minimal level of control in the process as the curve is narrow towards the minimum, but we did have a significant standard deviation which shows that there is tremendous scope of improvement, especially when the Coefficient of Variation is 1.20% for the first and 0.82% of the second. The difference indicates the interval between Arrival to Treatment is more variable than the interval between Treatment and Departure, indicating that major inefficiency is in the process from a patient's arrival to treatment. We will analyze this interval, henceforth known as 'treatmentInterval[1].

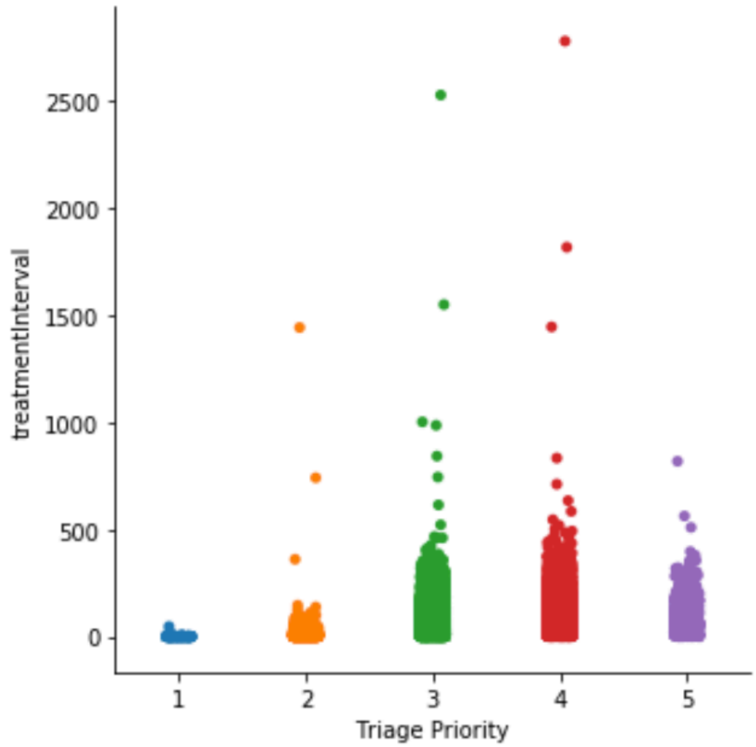
Naturally, the idea of priority arises, and it is plausible that patients with higher priority are treated faster than those with not. Hence, to better understand where our process lags, we need to look at this interval categorized by Triage Priority. We expect low triages to have a low time of treatment and extremely minimal standard deviation, because the more urgent the case, the more efficient and controlled the treatment process should be.

Interval Distribution by Triage Priorities

We plotted the treatInterval categorized by Triage Priority to analyze whether any of the sub-processes of dealing with specific triage priorities is poorly controlled.

---

[1] I understand that ideally we should not reduce the granularity/resolution of the data. It is highly possible that by taking the difference between the two columns, I am also compounding the effects of the noise within the column data. However, due to lack of time, I am just showing one particular stream of thought.



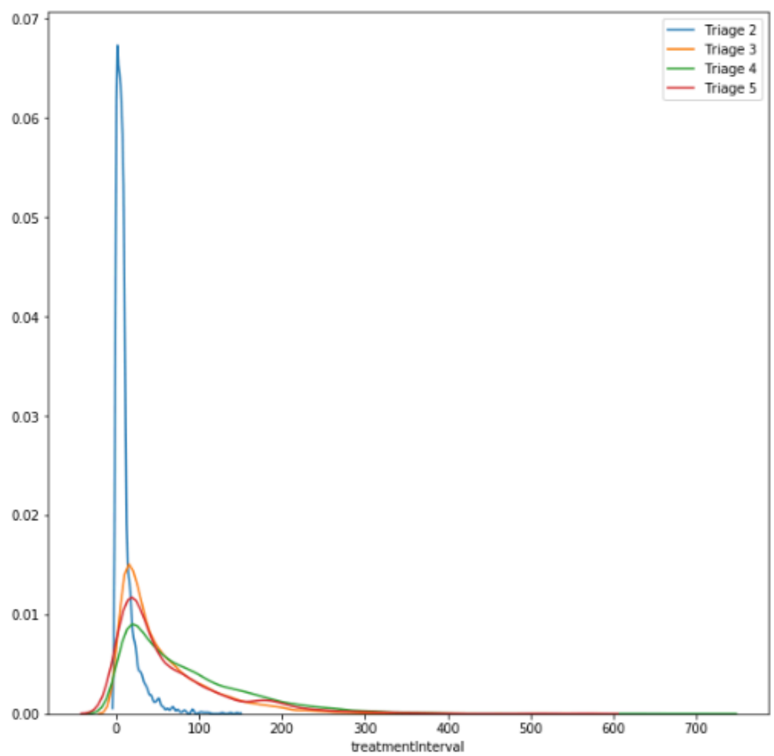
The scatterplot above shows high standard deviation value for triage priority categories 3,4 and to some extent 5, while low standard deviations for triage priority categories 1 and 2. This implies that the process of dealing with patients coming in with priorities 1 and 2 is well defined, well controlled, and consequently efficient in terms of the time taken by these patients to see the doctor/nurse. This inference is consistent with ideal since the interval between arrival and treatment for higher priorities will be smaller.

Meanwhile, the processes of dealing with patients with triage priorities identified as 3, 4, 5 had a high standard deviation and therefore are not well controlled, because of the high standard deviation.

It is perhaps valuable to explore these categories specifically and understand what particular variable from data correlate with triage values and see how they affect the treatmentInterval. Based on these insights, a decision maker may arguably focus on revisiting the processes to deal with triage priorities 3,4, and 5 and allocate more resources for their optimization.

We also see significant outliers for each case. These outliers may be relevant or irrelevant, but for purely normalized statistical relationships within our categories, we will risk dropping them[2]. We will bound our data to remove outliers to get more appropriate statistical inferences of how efficiently the patients in a particular triage category will be treated.

On plotting our cleaned data, we get:



Triage Priority	mean	std
2	11.154380	36.969349
3	57.348925	61.204023
4	83.396025	84.466497
5	65.075540	73.260135

Table 1: Mean and Std. Dev for Triage Priorities

A high-level insight is perceivable to the decision maker: most of the inefficiencies lie towards the lower triages, implying that a number of patients are coming in with low priority complaints, and that the process to deal with these patients takes more time. This is perhaps due to a lack of resources specifically for low triage complaints or simply a greater queue priority to high triages. Either way, if the organization overall aims to cut down on it average patient time and the variance in it, it will have to focus on optimizing these low triage priority cases.

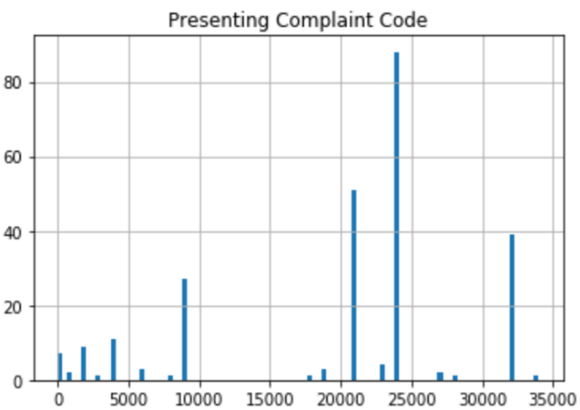
Correlation between Triage Priorities and Patient Complaints

The first factor which plausibly affects triage evaluation is the complaint the patient has.

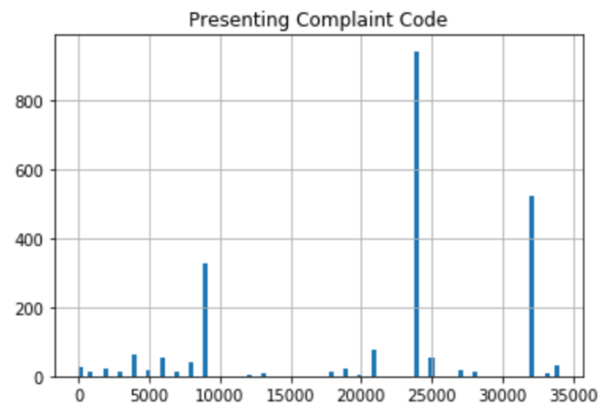
---

[2] Ideally, thorough analysis should be done on outliers in order to determine that are they extremely useful if relevant or simply errors. Considerable insight can be provided by outliers in terms of anomalies in processes or fracture points in the setup.

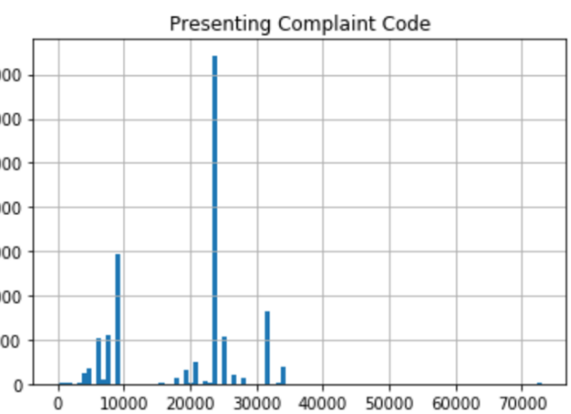
Although there were numerous, and probably more effective ways to find the relationship between the two, I simply plotted histograms to see which kind of complaints were most common in a triage category. This would give some indication towards which kinds to treatment were slower, which specific branches of the medical staff faced more patients, and therefore which specific areas would resources be allocated to, or would rank higher to cause overall process efficiency improvement.



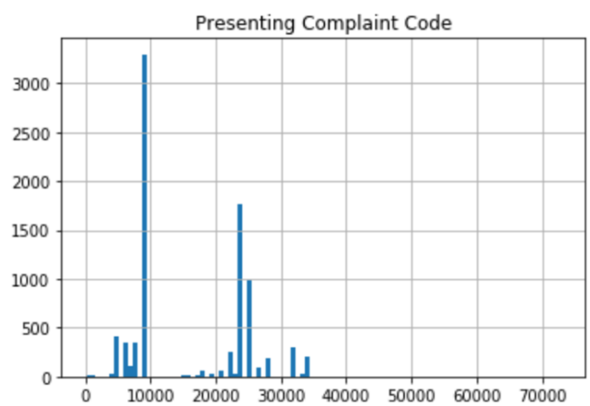
Triage Priority 1



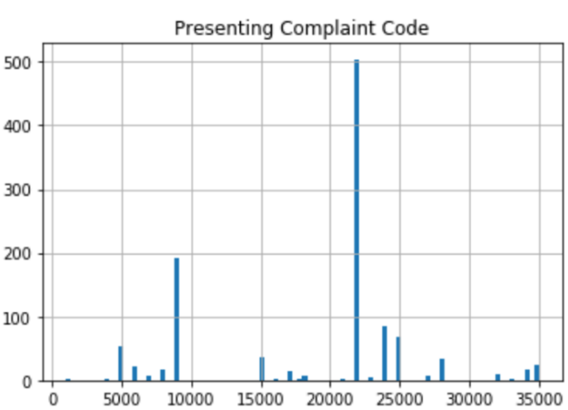
Triage Priority 2



Triage Priority 3



Triage Priority 4



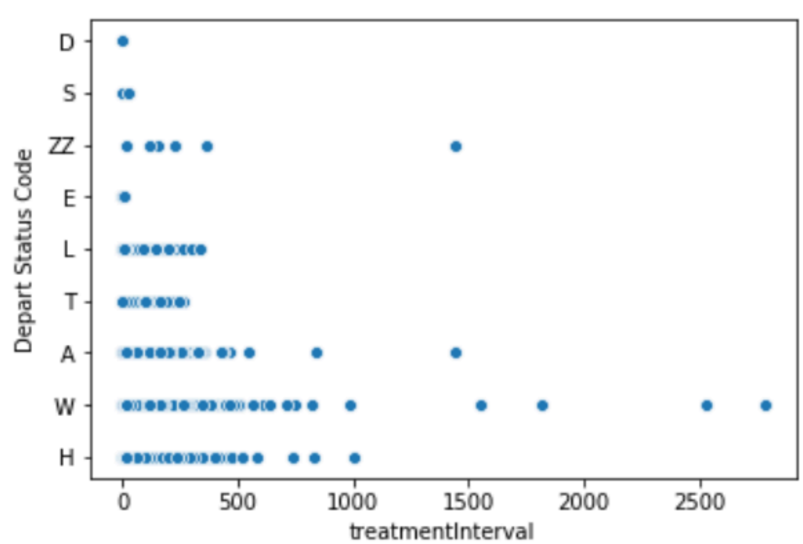
Triage Priority 5

Triage	1st Highest Comp. Code	2nd Highest	3rd Highest
1	24000 (88)	21000 (44)	32000 (32)
2	24000 (942)	32000 (468)	9000 (292)
3	24000 (7425)	9000 (2467)	32000 (1407)
4	9000 (2996)	24000 (1757)	25000 (888)
5	22000 (503)	9000 (170)	24000 (84)

Clearly, we see certain complain codes occurring most often than the other. It should be noted that 24000 occurs most frequently because our dataset represents samples aged between 0-15 with a mode of 0. However, 9000 and 32000 pertain to INJURY and RESPIRATORY respectively, suggesting more allocation of personnel equipment to reduce the interval between a patient's arrival and treatment.

Implications of Current Processes/Urgency of Improvement

I also looked at the relationship between treatment delay and the state in which patients departed as an overall indicator of the effectiveness of the current process, I simply plotted a scatterplot which categorically evaluated the distribution of departure with respect to treatmentInterval:



The plot showed that as treatmentInterval increased, the probability of patient not waiting increases. At the same time, those who left accounted for approximately 10.9% denoting a moderate sense of urgency to process improvement. 75.7% left healthy and 11% were admitted, giving the overall picture that the process was sufficing.

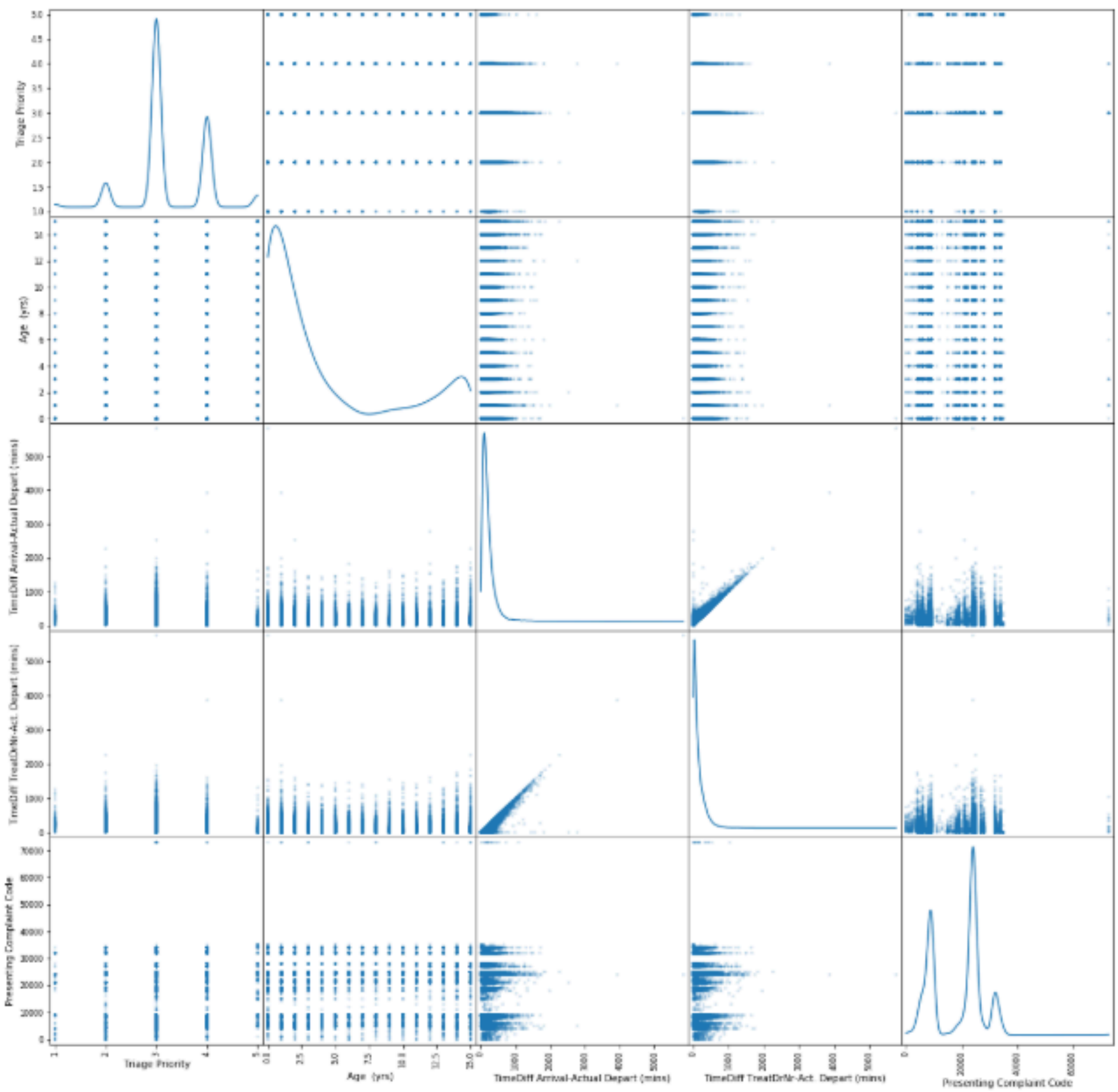
Conclusion

In conclusion, I perform a very narrow analysis focusing on evaluating whether the process of providing service to patient at an emergency department was efficient with respect to time. I explored in which specific categories of triage priorities did the inefficiencies lie and how were they linked to specific complaints, thus indicating specific target domains where a decision maker could focus on to achieve maximum overall improvement. I also performed a brief evaluation of how urgently the process is needed and whether the current system suffices.

Appendix A: Summary Analysis

Before data was analyzed, a quick summary analysis was performed to gain some insight on the raw data given.

A matrix of scatterplots between the different numerical variables for each sample was plotted to get an insight into the relationships within the data. The diagonal in the matrix



was modified to represent kernel density functions of the intersecting variable based on the Gaussian kernel.
The matrix helped infer the following relevant to our questions:

1. Majority of the samples had a Triage Priority of 3 and 4.

2. Majority of the samples had an age between 0-5 inclusive. The age had linear correlation with the time spent at the ED.

3. The Triage Priority had some form of relationship with both the Time Difference variables. Triage Priorities at the median and closer to it had higher waiting times compared to Priorities beyond the quartiles.

4. The two Time Difference variables had an obvious linear correlation, implying that 'Time Difference between Arrival and Departure' was proportional to the Time Difference between Treatment and Departure. This help introduces a plausible assumption that the bottleneck was the treatment. Thus, it would be valuable to evaluate the duration between a patient's Arrival and Treatment.