Patient Flow Modeling and Optimization in Hospital Emergency Departments with Discrete Event Simulations

Beatrice Bonato

May 9, 2024

ABSTRACT

This study presents a discrete event simulation (DES) model developed using the SimPy library to model, analyze and enhance patient flow within an emergency department (ED) of a U.S. hospital. Recognizing the critical need for efficient healthcare delivery, especially in busy emergency settings, the model incorporates real-world data on patient arrivals and distributions to simulate daily operations accurately.

The DES approach enables the identification of operational bottlenecks and the evaluation of various strategies to optimize resource allocation and patient throughput. By configuring different scenarios, this project highlights the key service areas where improvements can be made, thus improving the overall efficiency of the ED.

Key performance measures such as average waiting times, queue lengths, and service utilization rates across essential services: triage, doctor consultations, diagnostics, treatments, discharges, and admissions, were analyzed. These metrics are very important for understanding the dynamics of patient flow and for implementing effective management strategies.

The results from this simulation provide actionable insights that can improve patient management and satisfaction by reducing wait times and ensuring that critical resources are available when needed. This study demonstrates the practical applications of discrete event simulation technologies in healthcare settings, offering a tool for future enhancements in emergency department operations.

INTRODUCTION

The management of patient flow within healthcare facilities is a critical issue that significantly influences the quality of patient care, efficiency of healthcare delivery, and overall patient satisfaction. The complexity and dynamic nature of this environment, particularly in busy settings like emergency departments, necessitate sophisticated management strategies to ensure that healthcare services are both efficient and effective.

Efficient patient flow is crucial in healthcare because it directly affects treatment times, patient satisfaction, and the utilization of healthcare resources. Ineffective flow can lead to long waiting times, increased stress for both patients and healthcare personnel, and can even impact the quality of care delivered. In the United States, for example, overcrowding and long wait times in emergency departments have been associated with increased mortality rates, longer hospital stays, and higher costs [1][2]. These issues underscore the need for efficient patient flow management systems that can adapt to varying patient volumes and resource availability.

In Boston, extended wait times in emergency departments (EDs) have become a significant concern. Reports indicate that patients, including those with less critical health issues, often undergo delays going from 12 to 24 hours at some of the city's most prominent hospitals. For instance, at Massachusetts General Hospital and Brigham and Women's Hospital, 20% and 16% of patients, respectively, experienced ED stays exceeding 12 hours in 2022. These long waits, not only compromise patient safety but also strain hospital resources. Such scenarios underscore the urgent need for effective patient flow management strategies within healthcare systems, particularly in busy urban settings like Boston [8].

The current state of the art in patient flow management involves using various methodologies to model, analyze, and improve the flow of patients through healthcare facilities. Discrete Event Simulation (DES) has emerged as a particularly effective tool in this area. DES allows healthcare managers to simulate different operational scenarios and their effects on patient flow, enabling them to identify potential improvements in processes. This approach has been effectively applied in various healthcare settings across the United States, where it has helped to reduce waiting times and improve service delivery in hospitals [1][5].

Several studies have highlighted the effectiveness of using DES in healthcare settings. For example, a review of DES applications in healthcare showed that this technique is increasingly being used to address complex logistical challenges in patient management, often leading to significant operational improvements [1]. Specific case studies, such as those conducted at Indiana University Health Arnett Hospital and other healthcare facilities, have demonstrated how DES can optimize the scheduling of appointments in outpatient clinics, thus reducing waiting times and improving patient throughput [6]. Furthermore, in emergency departments, DES has been utilized to model patient arrivals, treatment, and discharge processes, providing valuable insights that have led to decreased patient waiting times and enhanced capacity planning [5].

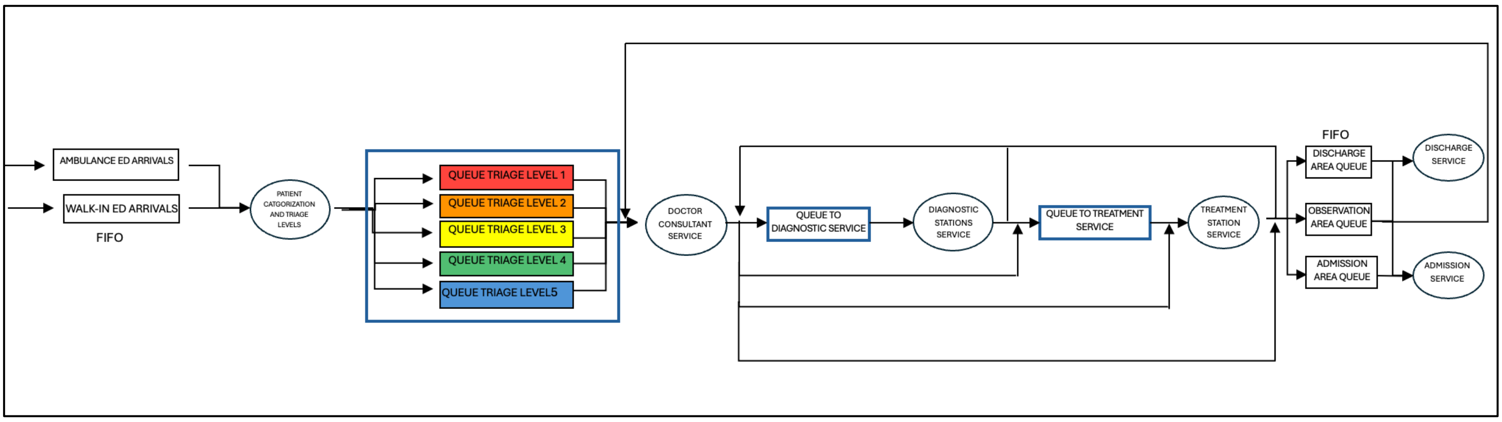
In conclusion, optimizing patient flow in healthcare settings is a complex but essential task that requires sophisticated modeling techniques and comprehensive analyses. By continuing to develop and apply these methodologies, healthcare providers in the United States and all over the World can significantly improve their operations and the level of care they are able to offer.

In response to these challenges, this project developed a discrete event simulation (DES) model using the SimPy library to simulate patient flow within an emergency department (ED). The model leverages empirical data on patient arrivals and patient distribution to provide a realistic representation of daily operations in a typical U.S. emergency department. By simulating various scenarios and parameter configurations, the project identifies bottlenecks and evaluates strategies to optimize patient flow and resource allocation. The main achievements of this project include the ability to accurately represent real-world scenarios, facilitating the identification and analysis of bottlenecks. Through the simulation, I was able to define and assess critical key performance measures such as average waiting times and average queue lengths throughout the day for major ED services including triage, doctor consultations, diagnostics, treatments, discharges, and admissions. Additionally, the project calculated the service utilization rates for each of these services. These performance measures are all crucial as they show the operational dynamics within the ED, allowing for targeted interventions to improve efficiency and patient care. Furthermore, the insights gained from this simulation offer a valuable foundation for future enhancements in ED management. By improving and applying the findings from this study, healthcare facilities can better manage patient volumes, minimize wait times, and allocate resources more effectively, leading to improved patient outcomes and satisfaction. This approach not only addresses the immediate needs of the emergency department but also contributes to broader healthcare system improvements, demonstrating the potential of simulation tools in complex healthcare environments.

METHODS

Methodology:

In this project, I developed a discrete event simulation (DES) model using the SimPy library in a Google Colab environment to accurately simulate the patient flow within an emergency department (ED). The model's structure is based on a detailed scheme that delineates various processes from patient arrival to discharge or admission, capturing the complexity and dynamics of an actual ED environment.



Model Structure and Assumptions:

The simulation begins with patient arrivals handled on a first-come, first-served basis. For simplicity we didn't make differences between patient arriving via ambulance and walk-ins' patients, even though we know from real data that on average 32.9% (2.2 standard deviation) of patients arrive on ambulance [9]. Every patient undergoes a triage assessment with a set service time of five minutes. Triage assessment in an emergency department (ED) is a critical process used to evaluate and prioritize patients based on the urgency of their medical needs. The goal of triage is to ensure that patients receive the most appropriate care at the correct time. Most EDs use a standardized triage scale, such as the Emergency Severity Index (ESI), which typically divides patients into five categories:

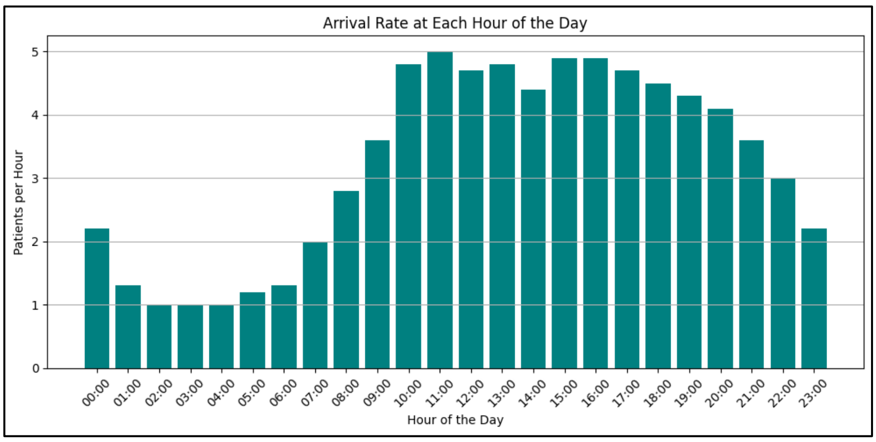
- Level 1: Immediate life-saving intervention required.
- Level 2: Very urgent, potentially life-threatening.
- Level 3: Urgent, but not immediately life-threatening.
- Level 4: Less urgent.
- Level 5: Non-urgent.

Therefore, during triage, patients are assigned a severity level from 1 to 5 which determines their priority in subsequent queues. After triage, all patients enter a priority queue for a doctor's consultation, where the most severe cases are treated first, with a fixed consultation time of 20 minutes.

Post-consultation, the model assumes that patients proceed to either diagnostics, treatment, or both, before being discharged or admitted. For simplicity, the simulation allows only one round of diagnostics or treatment per patient. Service times are set at 20 minutes for diagnostics, 40 minutes for treatment, 20 minutes for discharge, and 30 minutes for admission. These times were personally assumed based on general knowledge and practical considerations, recognizing potential deviations from real world conditions.

Data and Parameterization:

The simulation uses empirical data to model patient arrivals and service needs accurately. Patient arrivals are modeled using Poisson distributions with hourly rates derived from literature, reflecting realistic daily fluctuations/patterns. To get a more realistic DES model, I integrated empirical data on patient arrivals from a real US emergency department. This data is critical as it captures the hourly variability in patient influx, which is essential for modeling the dynamic environment of an ED. Specifically, the hourly arrival rates used in my simulation are [2.2, 1.3, 1.1, 1, 1, 1.2, 1.3, 2, 2.8, 3.6, 4.8, 5, 4.7, 4.8, 4.4, 4.9, 4.9, 4.7, 4.5, 4.3, 4.1, 3.6, 3, 2.2] patients per hour, beginning at midnight and proceeding through each hour of the day. This pattern highlights significant fluctuations, with patient arrivals peaking during midday and early evening hours [10].



Incorporating these real-world data points significantly enhances the accuracy of the simulation. It allows to realistically portray the daily operational challenges that emergency departments face. provides a solid basis for assessing patient flow and optimizing resource allocation strategies.

The assignment of triage levels is based on average percentages from many literature sources, with the distribution [0.9%, 13.1%, 57%, 24.8%, 4.2%] across the five urgency levels from 1 to 5 [11].

Patient needs for diagnostics and treatments and admission were probabilistically determined using data from the National Hospital Ambulatory Medical Care Survey of 2021, with 79.1% (1.3) of patients undergoing diagnostics, 76.2% (1.2) undergoing medication therapy, 45.9% (1.5) undergoing one or more procedures and 13.1% (1.1) being admitted to hospital [9].

key assumptions for my model

1. Arrival Pattern: Patient arrivals at the ED are assumed to follow a Poisson distribution, with different rates for each hour of the day, reflecting real-life variability in patient arrivals.

2. Service Times: The service times for the services (triage, doctor consultation, diagnostics, treatment, discharge, and admission) are assumed to be exponentially distributed. Most of these service times were personally estimated by me due to the lack of consistent data available, making the system's behavior highly dependent on these parameters. Therefore, the model needs more accurate and empirical service time data to enhance its reliability and applicability.

3. Severity Level Assignment: Each patient is assigned a severity level from 1 to 5 during triage. This assignment dictates the order of service in subsequent queues. Once assigned, triage levels do not change over the length of staying of the patient in the ED.

4. Queue Management: The queues for doctor consultation, diagnostics, and treatment are priority queues based on triage severity level. On the other hand, the queues for discharge and admission services are managed on a first-come, first-served (FIFO) basis.

6. Single Service Round, no Feedback Loop: Patients undergo only one round of doctor consultation, diagnostics, and treatment. This is a simplification since patients may require multiple rounds of consultation or treatment.

7. Personnel Shift Changes: It is assumed that shift changes for hospital staff (nurses, doctors, diagnostic technicians) occur instantaneously without any idle or transition periods, ensuring no interruption in patient service.

8. Consistent Performance: The performance of all personnel is considered consistent throughout their shifts. There are no variations in speed of service due to fatigue or other factors.

9. Exclusion of Observation Area: The model does not include an observation area/queue where patients might be monitored for changes before further decisions are made.

10. Infinite Hospital Beds and Seats: The model assumes an unlimited number of hospital beds for admission and an unlimited number of seats in the emergency department (ED). This is a big simplification.

Simulation Configuration and Key Performance Indicators (KPIs):

Various resource configurations were tested, such as different numbers of triage nurses, doctors, and diagnostic stations, to identify realistic setups. Key performance indicators measured included average and maximum wait times for each service area, resource utilization rates, and total patients processed.

This methodology allows to assess the efficiency and effectiveness of patient flow within an ED. By leveraging real-world data and a well-structured simulation process, the project offers valuable insights into operational dynamics and potential areas for improvement, crucial for enhancing patient care and resource management in emergency healthcare settings.

RESULTS

The simulation was conducted to understand the impact of resource allocation on patient flow within an Emergency Department (ED) using different sets of resources. Results for 2 of all the set of resources tried are provided below. Service times were set as follows across both simulations:

- Triage assessment: 5 minutes
- Doctor visit: 20 minutes
- Diagnostic service: 20 minutes
- Treatment service: 40 minutes
- Discharge service: 20 minutes
- Admission service: 30 minutes

Resource Set 1:

- 1 triage nurse, 2 doctors, 2 diagnostic stations, 3 treatment stations, 1 discharge station, and 1 admission station.

Resource Set 2:

- Maintained 1 triage nurse and 2 doctors but increased to 3 diagnostic stations and 2 discharge stations while keeping treatment and admission stations unchanged.

The differences between the two resource sets illustrate the significant impact that even small adjustments in resource allocation can have on patient waiting times and overall ED throughput. Specifically, the addition of one diagnostic station and one more discharge station in Resource

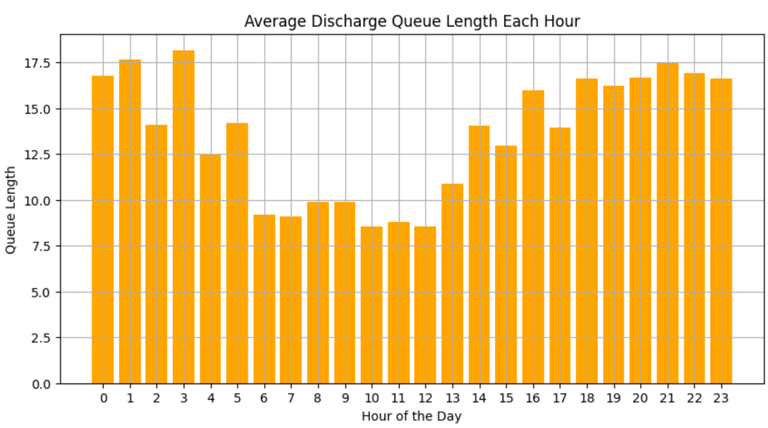
Set 2 led to marked improvements in wait times for these services, underlining the importance of resource distribution in managing patient flow.

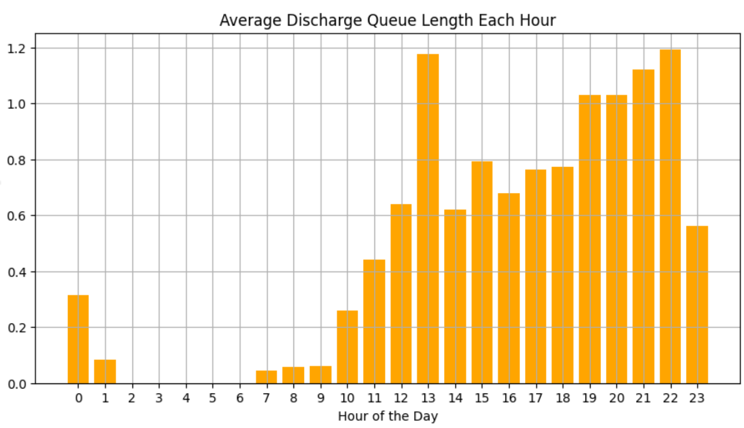
By examining the bar plots of queue lengths and hourly waiting times (provided in the Google Colab notebook), we can identify where bottlenecks typically occur. There is a noticeable pattern of increased waiting times during peak hours from around 10 AM to 8 PM. This suggests that strategic allocation of additional resources during these times could effectively reduce bottlenecks and optimize patient flow.

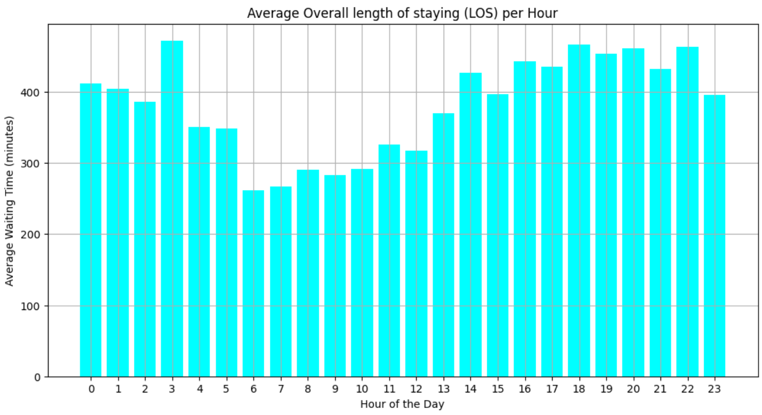
The performance measures obtained from the simulation is crucial for making smart decisions about resource management in an ED. The substantial reduction in waiting times for diagnostics and discharge services in Resource Set 2 highlights how targeted adjustments can alleviate specific bottlenecks. This approach not only improves patient satisfaction by reducing their length of stay but also enhances the overall efficiency of hospital operations. Results for Resource set 1 are on the left while results for Rehouse set 2 are on the right.

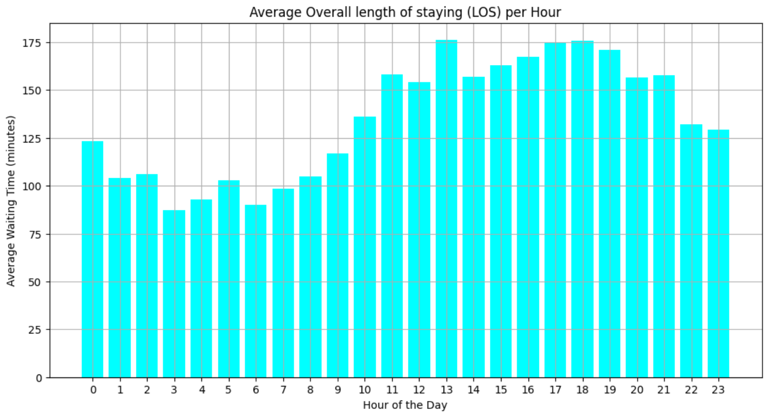
KEY PERFORMANCE MEASURES SET 1	
Average Wait Time for Doctor Service	31.26 minutes
Maximum Wait Time for Doctor Service:	453.93 minutes
Average Wait Time for Diagnostic Service:	7.15 minutes
Maximum Wait Time for Diagnostic Service	175.18 minutes
Average Wait Time for Treatment Service:	18.64 minutes
Maximum Wait Time for Treatment Service	853.18 minutes
Average Wait Time for Discharge Service	283.48 minutes
Maximum Wait Time for Discharge Service	876.05 minutes
Average Wait Time for Admission Service	7.89 minutes
Maximum Wait Time for Admission Service:	145.91 minutes
Utilization Rate for Triage	26.93%
Utilization Rate for Doctor	52.53%
Utilization Rate for Diagnostic	41.09%
Utilization Rate for Treatment	54.46%
Utilization Rate for Discharge	90.93%
Utilization Rate for Admission	21.30%
Average Length of Stay	386.38 minutes
Total patients processed over 30 days	2303

KEY PERFORMANCE MEASURES SET 2	
Average Wait Time for Doctor Service	35.64 minutes
Maximum Wait Time for Doctor Service:	475.55 minutes
Average Wait Time for Diagnostic Service:	1.55 minutes
Maximum Wait Time for Diagnostic Service	55.87 minutes
Average Wait Time for Treatment Service:	19.35 minutes
Maximum Wait Time for Treatment Service	407.64 minutes
Average Wait Time for Discharge Service	10.22 minutes
Maximum Wait Time for Discharge Service	124.90 minutes
Average Wait Time for Admission Service	8.99 minutes
Maximum Wait Time for Admission Service:	116.74 minutes
Utilization Rate for Triage	26.90%
Utilization Rate for Doctor	51.73%
Utilization Rate for Diagnostic	29.00%
Utilization Rate for Treatment	54.19%
Utilization Rate for Discharge	45.36%
Utilization Rate for Admission	19.63%
Average Length of Stay	147.79 minutes
Total patients processed over 30 days	2331









The results clearly indicate a significant bottleneck in the discharge service within the first resource set, which substantially contributes to the overall Length of Stay (LOS) for patients. It's interesting to observe that merely adding one additional resource to the discharge station dramatically reduces the waiting time for discharge from 283.48 minutes to just 10.22 minutes. This big improvement is also reflected in the queue length plots, where we can visually confirm the reduction in congestion at the discharge service station. This underscores the importance of resource allocation in managing patient flow efficiently in emergency departments.

Resource allocation Optimization

In my simulation, I then incorporated the capability to dynamically adjust resources based on observed needs, which were indicated by the queue lengths at each service point. The function `adjust_capacity` in my code continuously monitors the total queue length for each resource every hour. When the queue length surpasses a predefined threshold and additional resources are available, the function increases the number of active resources. Conversely, if the queue length is significantly low, indicating underutilization, it reduces the number of resources to prevent them from sitting idle.

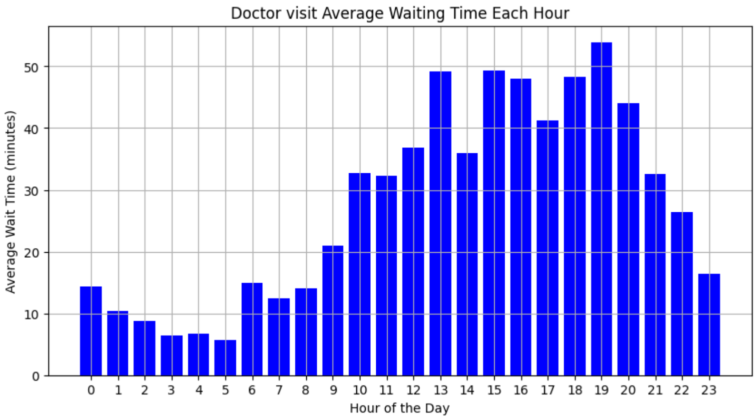
By dynamically adjusting the resources, I observed some improved results overall. From the waiting time plots for doctor visits and treatment below, it's evident that there is a better distribution of resources. This can be seen from the fact that the plots are less skewed and more symmetric compared to when resources are statically allocated. These improvements can be seen from a comparison between a 30-day simulation using the static resource of Set 2 and another simulation employing dynamic resource allocation, but still starting with the same resources as in set 2. The configuration for this simulation was as follows:

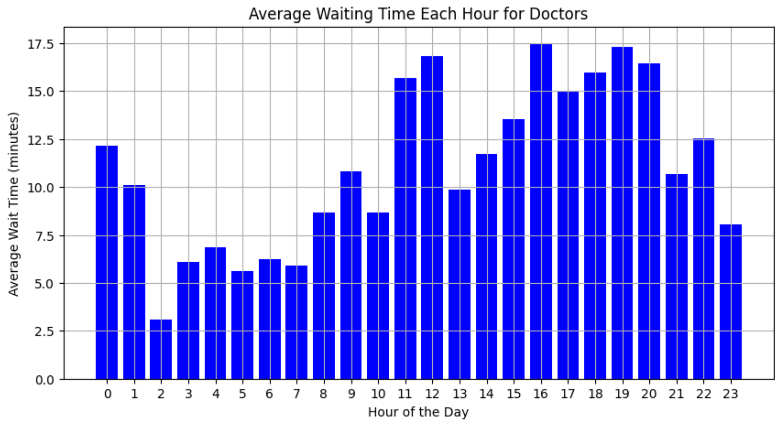
```
resources = {
    'triage': DynamicResource(env, 1, 10, 1),
    'doctors': DynamicResource(env, 2, 10, 1),
    'diagnostic': DynamicResource(env, 3, 10, 1),
    'treatment': DynamicResource(env, 3, 10, 1),
    'discharge': DynamicResource(env, 2, 10, 1),
    'admission': DynamicResource(env, 1, 10, 1) }
```

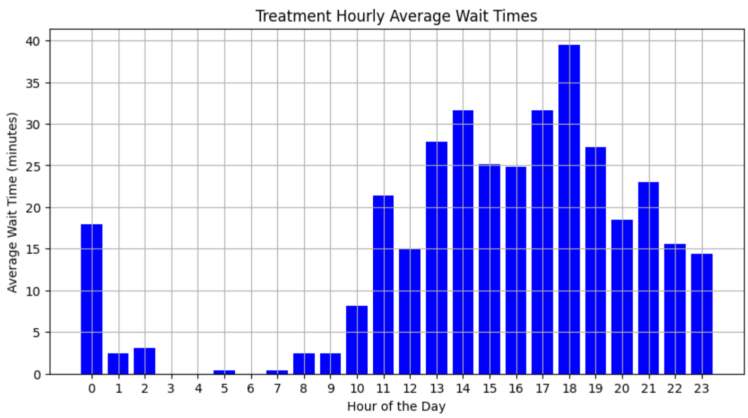
Each `DynamicResource` has three key values:

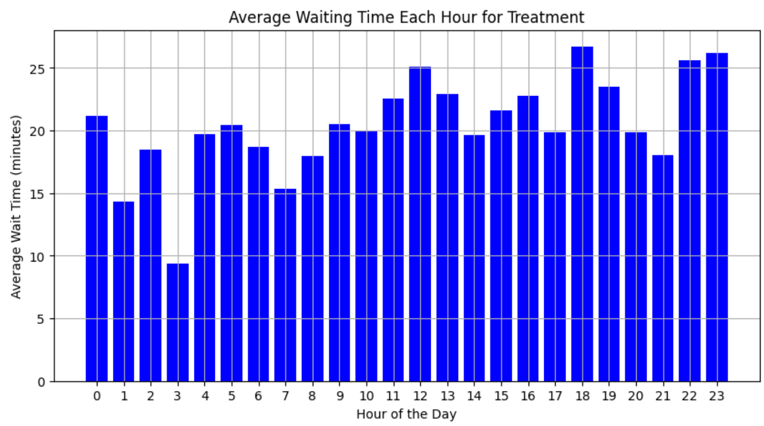
- Initial Capacity: The starting number of units available (e.g., 2 doctors).
- Maximum Capacity: The maximum number of units that can be activated (e.g., up to 10 doctors).
- Threshold: A queue length that triggers adjustments (e.g., adding resources if the queue exceeds this value).

The `adjust_capacity` function dynamically modified the resources as follows: if the queue lengths exceeded a set threshold, it added 3 more resources; conversely, if the queue length was smaller than one and there were more than 5 active resources, it removed one resource. I opted for more aggressive thresholds and conditions because using smaller thresholds did not yield significant changes and tended to create unnecessary fluctuations. Results for Static Resources allocation are on the left while results for Dynamic Resources allocation are on the right.









CONCLUSION:

The project successfully developed a simulation model using the SimPy library to represent an emergency department (ED) accurately. This model was then enhanced by dynamically adjusting resources based on real-time demand, achieving improvements in patient flow and resource utilization. Such advancements have reduced waiting times and optimized the distribution of workloads, enhancing both operational efficiency and patient satisfaction. The ability to accurately represent an ED system is extremely important, providing a useful tool for exploring various scenarios and testing different resource management strategies without disrupting real-world operations.

Incorporating empirical data from existing literature was a critical step in ensuring the realism of the simulation. This data allowed for the establishment of baseline conditions that reflect actual ED operations, making the insights gained from the simulation possibly relevant for real-world applications. The simulation serves as a platform for decision-making and operational planning, allowing hospitals to evaluate the potential impact of different staffing levels, procedural changes, and other strategic adjustments.

Looking ahead, a valuable improvement on the project could be investigating in more detail the optimal thresholds for resource adjustments across various services within the simulation, each with differing service times. This analysis could help identify the most effective resource levels needed to handle fluctuating patient volumes without overloading or underutilizing hospital staff and facilities. Additionally, exploring the relationships between service times and patient outcomes could provide further insights into how adjustments in these thresholds might impact overall efficiency and patient satisfaction. This approach would not only refine the existing model but also enhance its applicability to a wider range of scenarios within the hospital setting.

By continuing to develop and refine this simulation model, healthcare facilities could improve their operational efficiency and quality of care. The simulation's ability to test various scenarios and its adaptability to real-time data make it a potentially useful tool in healthcare management, helping facilities to adapt to changing demands and improve patient outcomes effectively.

REFERENCES

[1] Jesús Isaac Vázquez-Serrano , Rodrigo E. Peimbert-García, and Leopoldo Eduardo Cárdenas-Barrón, Discrete-Event Simulation Modeling in Healthcare: A Comprehensive Review, Int. J. Environ. Res. Public Health 2021.

[2] Xiange Zhang, Application of discrete event simulation in health care: a systematic review, Zhang BMC Health Services Research (2018)

[3] Sultanah Al Harbi ,1 Baker Aljohani,2 Lamiaa Elmasry,3 Frenk Lee Baldovino,1 Kamille Bianca Raviz,1 Lama Altowairqi,4 Seetah Alshlowi, Streamlining patient flow and enhancing operational efficiency through case management implementation, March 30, 2024

[4] Nathan Eddy, Using EHRs to track patients in real time, Healthcare IT News February 14 2020

[5] Nikkhah Ghamsari, Behnam, Modeling and Improving Patient flow at an Emergency Department in a Local Hospital Using Discrete Event Simulation, Thesis or Dissertation December 2017

[6] https://www.anylogic.com/resources/case-studies/outpatient-appointment-scheduling-using-discrete-event-simulation-modeling/

[7] Merrit Hawkins, Survey of Physician Appointment Wait Times and Medicare and Medicaid Acceptance Rates, 2022

[8] https://www.bostonglobe.com/2024/02/05/metro/massachusetts-er-patients-wait-for-care/

[9] NATIONAL CENTER FOR HEALTH STATISTICS, National Hospital Ambulatory Medical Care Survey: 2021 Emergency Department Summary Tables

[10] Naser M. Chowdhury, PE, LSSBB | Lawrence Riddles, MD, MBA, FACS, FACPE | Richard Mackenzie, MD, MBOE, LSBB, FACEP, Using Queuing Theory to Reduce Wait, Stay in Emergency Department, September 14, 2018

[11]Qian Cheng a, Nilay Tanik Argon a, Christopher Scott Evans b,⁎, Yufeng Liu a,c,d,e,f, Timothy F. Platts-Mills g,h, Serhan Ziya a, Forecasting emergency department hourly occupancy using time series analysis, 25 April 2021