Università degli Studi di Milano
Dipartimento di Informatica
Course “Simulation” - Exam Project, A.Y. 2024-2025 Spring
Student: Yusuf Kemahlı
Project Title: Emergency Room Simulation

Introduction:

This project presents a discrete-event simulation model of an emergency room (ER). The model was developed to explore the performance and dynamics of a hospital system operating under uncertainty and with limited resources. The emergency room environment is characterized by the arrival of patients on an unpredictable basis, triage-based prioritization, variable treatment durations, and resource contention.

The model simulates the flow of patients from arrival to exit, passing through key stages such as triage assessment, doctor consultation, and optional laboratory testing. Upon arrival, patients are classified into one of three triage levels based on the severity of their condition: critical, moderate, or mild. Each category is associated with distinct service needs and behavioral traits. For instance, mild patients are more likely to leave due to impatience, while critical patients receive the highest service priority.

The objective of this simulation is to evaluate how factors such as doctor availability, patient prioritization, and waiting tolerance influence key system metrics. The model also supports scenario-based experimentation, such as modifying resource configurations or timeout thresholds, to observe their effects on waiting time, patient loss, throughput, and resource utilization.

Model:

The simulation model represents the operational workflow of an emergency room. It uses a discrete event simulation approach to capture the time-driven interactions between patients, staff, and service units. The model has been developed to analyze system behavior under stochastic conditions, such as random patient arrivals, service durations, and patient-specific treatment paths.

The system boundaries are defined to include the patient journey from arrival at the emergency room to final exit, either after receiving care or leaving due to excessive waiting. To maintain the focus of the model and ensure computational efficiency, elements such as ambulance handling, surgical procedures, and detailed post-treatment care were intentionally excluded. These simplifications allow the model to concentrate on the most critical flow components while remaining extendable in future iterations.

The patient flow is structured as follows:
Arrival → Nurse Reception → Triage Queue → Doctor Consultation → [Laboratory Test if required] → [Second Doctor Visit] → Exit

At the time of arrival, each patient is assigned to one of three triage levels: critical, moderate, or mild. These categories are used to determine both the queue priority and the distributions used for service durations.

Service times were modeled using triangular or exponential distributions, as indicated by the latest literature and empirical studies. The probability of requiring a lab test was set to 30%. All parameters were selected to reflect real-world emergency room behavior, as supported by public healthcare data and academic research. For instance, patients' tolerance to waiting before leaving the system is modeled via a uniform random timeout, varying across patient types.

The emergency room model includes two main resource pools:

Doctors: with variable availability determined by a weekly shift schedule.

Nurses: statically available at the reception stage.

In order to maintain the authenticity of the simulation while ensuring scalability, we initially developed a model of a small-sized hospital, assuming an average of 100 patient arrivals per day. The number of doctors and nurses was determined to be the available resources. These parameters can be adjusted with ease to simulate different scenarios without altering the system structure, providing a flexible and reusable simulation framework.

Due to its event-based structure, the simulation tracks the status of all agents through time-triggered transitions. Patient priority logic, queue management, and resource contention are all handled dynamically through conditional routing and process logic.

While the current implementation focuses on key system constraints and performance factors (e.g., waiting times, utilization, and impatience), the structure of the model allows for straightforward extension. For example, the model can be extended to include more treatment types, patient conditions, or policy interventions.

Implementation:

The simulation model was implemented in AnyLogic using the Process Modeling Library, which provides high-level blocks suitable for discrete-event simulation. The main process flow is constructed within the Main agent using components such as Source, Service, Queue, SelectOutput, Delay, and Sink.

The model defines a custom Patient agent with the following attributes:

- type : triage level (Critical, Moderate, Mild).
- arrivalTime, exitTime : for total time tracking.
- triageWaitTime, queueEnterTime : for queue-specific KPIs.
- labRequired : whether the patient is sent for lab test.

The patient arrival process is modeled using an exponential interarrival time distribution, a common choice for emergency departments where arrivals are memoryless and irregular. Assuming an average of 100 patients per day, the mean interarrival time is 14.4 minutes, equivalent to a rate of $\lambda = 0.0694$.

Following the creation process, patients are guided to the Reception Service block, which is managed by a Nurse Resource Pool. This stage, referred to as the triage assignment stage, has a mean duration of five minutes and is modeled using an exponential distribution, consistent with literature on typical triage durations.

Each patient is then assigned a triage level using a uniform random number generator, mapped to the following probabilities:

- Critical: 20%
- Moderate: 50%
- Mild: 30%

These proportions are based on international emergency department statistics and ensure a realistic mix of severity levels.

The patient flow is then directed to a priority-based queue, where patients are organized according to their triage level (Critical > Moderate > Mild). At this stage, timeout behavior is modeled using the Enable exit on timeout feature. The timeout limit for each patient is set using a uniform distribution (e.g., Mild patients: Uniform(75,120)). These values were selected based on studies indicating that low-priority patients in overcrowded systems frequently depart after waiting 60–120 minutes.

To manage demand and ensure efficient allocation of resources, hold blocks are utilized before consultation blocks. This approach ensures that patients are seen only when doctor resources are available, improving the flow of patients through the system.

Medical consultations are managed through Service blocks connected to a Doctor ResourcePool, whose availability is defined by a daily schedule (e.g., 1 doctor from 00:00–08:00, 2 doctors from 08:00–16:00, 1 doctor after 16:00). This scheduling structure aligns with standard hospital staffing policies and allows for flexible adjustments across time periods.

The duration of consultations is modeled using exponential distributions with patient-type-based means:

- Critical: Exponential distribution with mean 20
- Moderate: Exponential distribution with mean 15
- Mild: Exponential distribution with mean 10

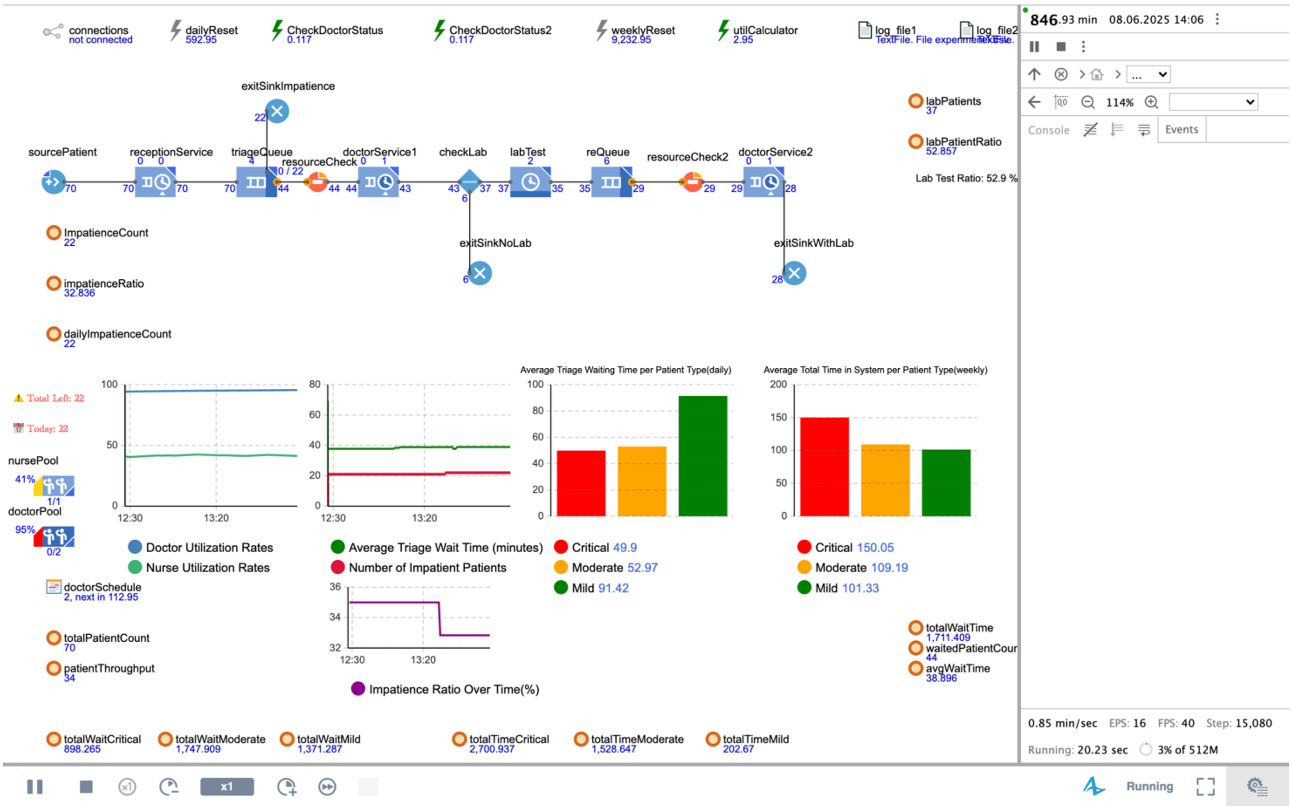
Following the initial doctor consultation, patients are assigned a 30% probability of requiring a laboratory test, modeled by evaluating uniform(0,1) < 0.3. The value is stored in labRequired and processed through a SelectOutput block. If no test is required, the patient exits directly; otherwise, they are routed to the Lab Test Delay, with service time drawn from a triangular distribution (min = 15, max = 60, mode = 30). This distribution was selected to reflect the variability and typical processing times of common emergency lab tests.

Following the completion of the laboratory test, patients are re-directed to a second priority queue and then proceed to another service block (DoctorService2) for a follow-up consultation. This block employs patient-type-dependent exponential service times, aligning with the earlier logic.

Finally, all treated patients exit via a Sink block. Individuals who surpass their designated time limit are able to exit through a dedicated path, facilitating KPI tracking on patient loss and waiting behavior.

KPI data is collected using On Enter and On Exit blocks. This process updates global variables, including average wait time, total time in system, patient throughput, lab test ratio, and impatience counts. These metrics are best represented using Text displays, Time plots, and Bar charts. The real-time visuals facilitate debugging and provide valuable feedback during scenario experiments.

Some model variables are reset periodically via cyclic event blocks, allowing daily or weekly analysis. The model's architecture allows for flexibility in modifying core variables , such as patient arrival rate, number of staff, and timeout thresholds , without requiring structural changes, enabling rapid experimentation and comparative analysis.



Runtime snapshot of the Emergency Room Simulation model in AnyLogic, displaying patient flow, resource utilization, and live KPI monitoring.

Key System Parameters (KSP) and Key Performance Indicators (KPI):

In order to evaluate the behavior and performance of the Emergency Room Simulation, several input parameters (KSP) were identified as critical to the system's dynamics. Similarly, a set of well-defined performance indicators (KPIs) were selected to assess the overall outcomes of the simulation, especially in terms of responsiveness, efficiency, and service quality.

(a) Key System Parameters (KSP)

The following parameters were identified as having a significant impact on the system's behavior:

- **Number of doctors:**
  This has a direct impact on system capacity and service rate. Increasing doctor availability has been shown to reduce waiting time and patient congestion. However, it is important to note that excessive increases may lead to a decrease in utilization.
- **Interarrival rate (λ):**
  It is used to regulate the intensity of patient inflow. An increase in the arrival rate can lead to an increase in queue length, resource contention, and system stress.
- **Timeout thresholds per patient type:**
  It is essential to establish clear parameters regarding the maximum wait time before patients

consider leaving. Lower thresholds have been shown to lead to increased patient attrition and higher turnover; longer thresholds have been shown to increase pressure on doctors.

- **Probability of requiring lab tests:**
  This has a significant impact on the total time spent in the system and creates bottlenecks. Consequently, the implementation of additional laboratory tests has led to prolonged patient stays and the formation of secondary queues.

- **Triage distribution:**
  This tool is used to determine the proportion of patients in each urgency category. A skewed distribution (e.g., more critical patients) increases demand on high-priority resources.

These parameters are defined as global variables in the model and can be easily adjusted to test different scenarios without altering the structural logic.

(b) Key Performance Indicators (KPI)

The system's performance is evaluated using the following outcome measures:

- **Average triage waiting time (per patient type):**
  This KPI indicates the average wait time for patients to be seen by a doctor. This is a key indicator of responsiveness.

- **Total time in system (per patient type):**
  It reflects the entire patient journey and system efficiency.

- **Patient throughput:**
  This is a metric used to gauge the efficiency of healthcare providers, measuring the number of patients who are successfully treated and discharged within a specified timeframe.

- **Number and percentage of impatient patients:**
  It captures patient dissatisfaction and highlights areas where service falls short of acceptable delay thresholds.

- **Doctor and nurse utilization rates:**
  This is an important metric for evaluating resource efficiency. This tool is designed to assist in the identification of staffing levels that exceed or fall below the optimal range.

These indicators facilitate a comprehensive evaluation of the emergency room's operational performance and provide actionable insights for potential improvements.

Experimental analysis:

In this section, we present an experimental analysis designed to assess how key system parameters (KSP) influence critical performance indicators (KPI) in the Emergency Room Simulation. Three structured experiments were conducted to explore the effects of:

1. Changes in doctor availability across daily shifts,
2. Variations in patient arrival rates, and
3. Different probabilities of patients requiring laboratory tests.

Each experiment was conducted over a 30-day simulation period and repeated 30 times per parameter configuration to ensure statistical reliability. The Parameter Variation experiment type in AnyLogic was used to systematically evaluate each configuration. During each run, key performance metrics such as average triage waiting time (per patient type), total time in system, patient throughput, impatience ratio, and doctor/nurse utilization were tracked.

The simulation outputs were exported to .csv files using AnyLogic's built-in TextFile elements. These data files were then processed using Python to generate visualizations, enabling clearer comparisons between scenarios.

The results of each experiment are presented below with relevant graphical analysis and interpretation.

Experiment 1: Doctor Shift Capacity

In this experiment, the focus was on evaluating how varying the number of available doctors during different daily shifts affects the overall performance of the emergency room system. The number of doctors directly impacts patient flow, waiting times, impatience behavior, and the effective utilization of resources.

Objective:
The objective of this study is to examine how enhancing doctor availability across different shifts affects patient waiting times, throughput, and resource utilization metrics.

Experimental Setup:
The experiment was conducted using two doctor shift configurations:

Scenario A (base case): The facility is staffed by one doctor on the morning shift, two doctors on the afternoon shift, and one doctor on the night shift.

Scenario B (Increased capacity): Two doctors are on the morning shift, three on the afternoon shift, and two on the night shift.

Each scenario was simulated over 30 days, with each day representing a full operating cycle (24 hours).

To ensure a controlled comparison, the patient arrival rate and triage distribution were kept constant across both scenarios.

Methodology:
A parameterized doctor schedule was implemented using AnyLogic's ResourcePool and Schedule blocks.

The simulations were executed in batch mode, and key performance indicators (KPIs) such as average triage waiting time (by patient type), total time spent in the system, patient throughput, percentage of impatient patients, and resource (doctor/nurse) utilization were recorded.

The simulation results were exported in CSV format and visualized using Python to produce bar charts and comparison plots.

Results & Observations:

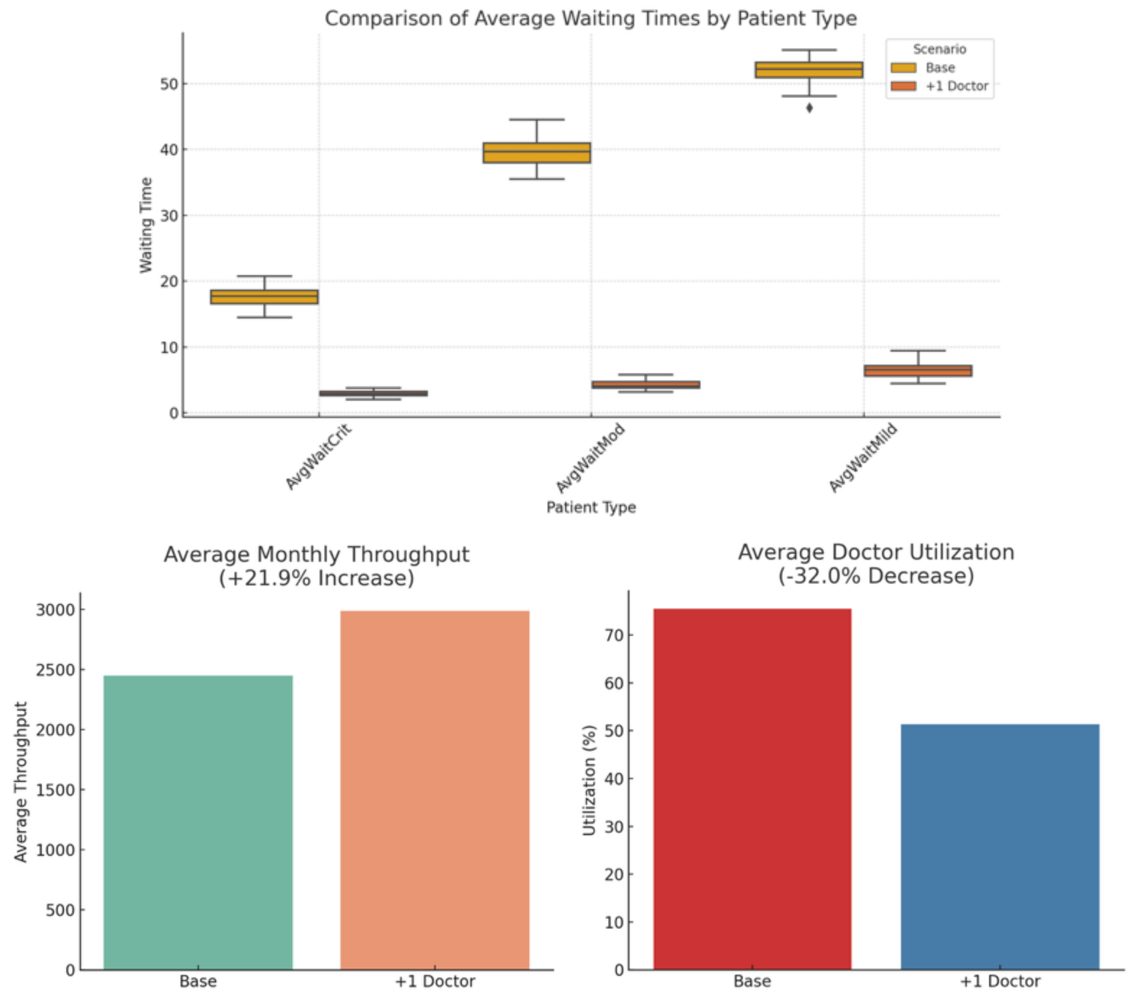
A substantial decrease in average waiting time was observed in Scenario B compared to Scenario A across all patient categories (critical, moderate, and mild).

The patient impatience ratio saw a significant decrease, indicating enhanced responsiveness and a reduction in patients abandoning the queue.

Doctor utilization decreased as capacity increased, indicating that while more doctors improved patient flow, their idle time increased slightly, suggesting a trade-off between availability and utilization.

Throughput improved notably in Scenario B, confirming that higher staffing levels result in increased service capacity.

Experimental Outcomes of Doctor Capacity Variation:



Experiment 2: Patient Arrival Rate

In this experiment, the aim was to evaluate how varying the number of patients arriving per day influences the overall performance and efficiency of the emergency room system. Patient inflow is a critical factor affecting service demand, queue lengths, waiting times, and resource strain.

Objective:

The objective of this study is to investigate how changes in daily patient arrival rates impact average waiting times, system congestion, resource utilization, and patient loss due to impatience.

Experimental Setup:

The experiment was conducted using four different patient frequency levels:

- **Scenario A:** ~72 patients/day ($\lambda \approx 0.05$)
- **Scenario B:** ~100 patients/day ($\lambda \approx 0.069$)
- **Scenario C:** ~144 patients/day ($\lambda \approx 0.1$)
- **Scenario D:** ~200 patients/day ($\lambda \approx 0.139$)

In all scenarios, the doctor schedule, triage distributions, and lab requirements were held constant to isolate the effect of increased patient inflow.

Each scenario was simulated over a 30-day period, with a full operating cycle (24 hours) per day, and repeated 30 times per configuration to ensure statistical reliability.

Methodology:

The inter-arrival times were modeled using an exponential distribution to reflect stochastic patient arrivals. Different arrival frequencies were implemented by adjusting the parameter λ of the exponential function.

AnyLogic's parameter variation experiment was used to automate the simulation runs. Key performance indicator (KPI) metrics such as average triage wait times (by patient type), total time in system, throughput, impatience ratio, and doctor/nurse utilization were captured.

All data from the simulations were written to CSV format and visualized using Python to produce boxplots and bar charts for comparative analysis.

Results & Observations:

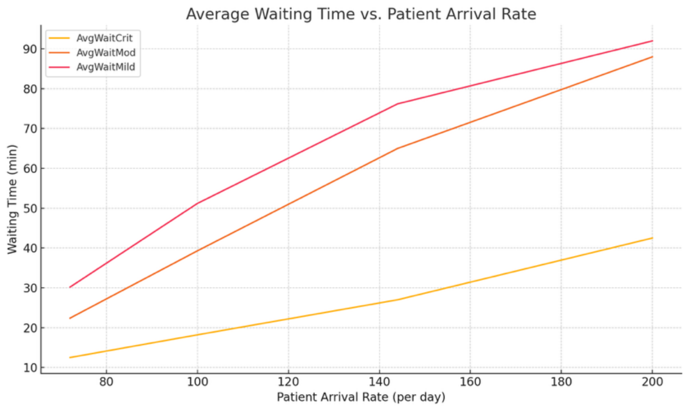
Waiting Times Increased with Load: As the number of arrivals increased, the average triage waiting times exhibited a marked increase, particularly in scenarios C and beyond. This suggests that the system's processing capacity was reaching its limits.

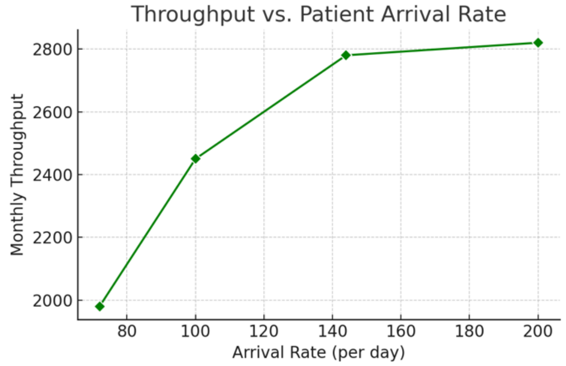
There has been a notable increase in the impatience ratio. The percentage of patients leaving the queue without service increased in proportion to the increase in arrival rates, indicating that the patient experience was negatively impacted by high loads.

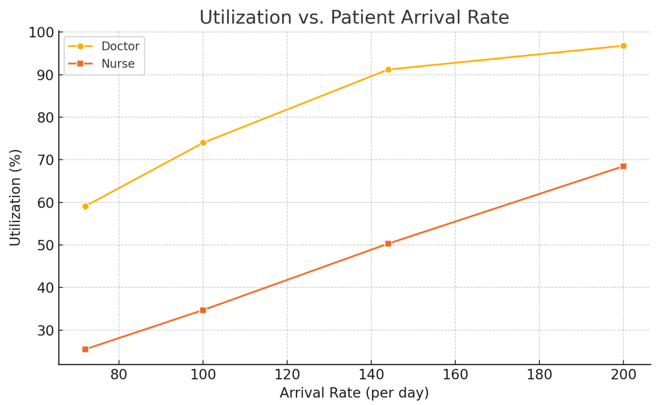
The utilization of resources has reached maximum capacity. There was a significant increase in the utilization of doctors and nurses due to rising demand, which reduced the system's flexibility and led to an increase in workload.

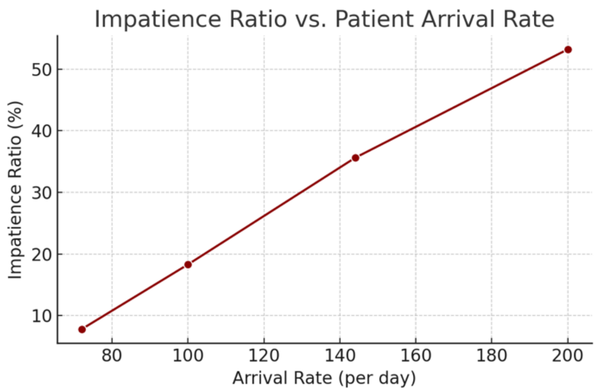
Throughput Plateaued: While more patients were processed in absolute numbers, throughput gains diminished in higher-load scenarios due to increased delays and patient loss.

Experimental Outcomes of Patient Arrival Rate Variation:









Experiment 3: Laboratory Test Requirement Rate

In this experiment, the aim was to evaluate how varying the percentage of patients requiring laboratory tests influences the overall performance and efficiency of the emergency room system. Diagnostic load is a critical factor affecting patient flow, treatment delays, and resource utilization. As the number of patients requiring lab testing increases, system bottlenecks may emerge, resulting in longer wait times and placing additional strain on clinical and support staff.

Objective:
The objective of this experiment was to examine how increasing the lab test requirement rate impacts key system performance indicators such as average triage waiting time, total time in system, patient throughput, percentage of impatient patients, and doctor/nurse utilization.

Experimental Setup:
The experiment was conducted by systematically adjusting the probability that a patient would require a lab test, while keeping all other system parameters constant. These parameters included the arrival rate, staffing, and triage distribution. Five distinct probability levels were evaluated in the study:

- Scenario A: 10% of patients require lab tests

- Scenario B: 30% of patients require lab tests
- Scenario C: 50% of patients require lab tests
- Scenario D: 70% of patients require lab tests
- Scenario E: 90% of patients require lab tests

Each scenario was simulated for 30 days, with 30 repetitions per configuration to account for stochastic variability.

Methodology:

A boolean variable, labRequired, was defined in each patient agent and assigned using a probabilistic condition (e.g., uniform(0,1) < 0.3 for a 30% lab requirement). The SelectOutput block directed patients based on this flag, sending those who required lab work to the Laboratory Test block before returning to the doctor queue. The model incorporates a triangular distribution to simulate the time required for laboratory testing, with the distribution based on realistic diagnostic durations.

Simulations were executed using parameter variation in AnyLogic. The resulting KPI values were then written to CSV files and visualized using Python through a series of comparison graphs.

Results & Observations:

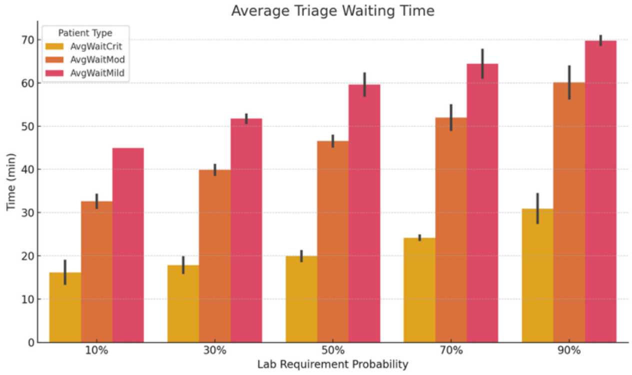
As the rate of lab test requirements increased, average waiting times and total time in the system rose significantly, especially for moderate and mild cases.

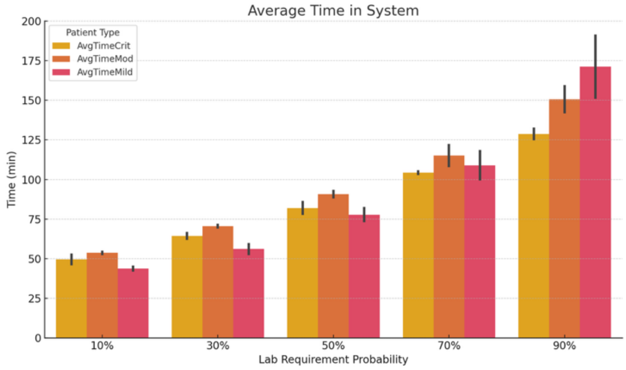
The impatience ratio has increased, indicating a higher rate of patient abandonment due to prolonged delays.

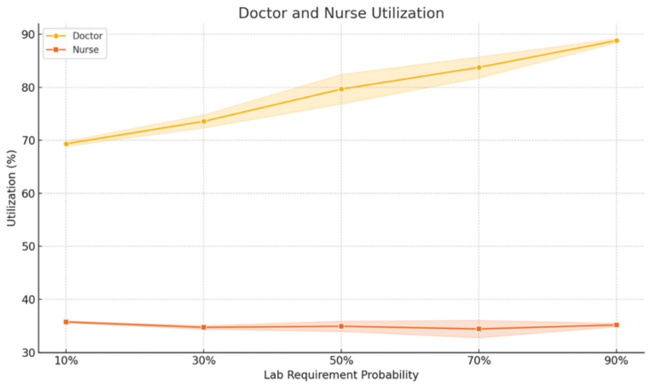
Doctor utilization increased in response to longer care cycles and diagnostic feedback loops, while nurse utilization remained largely unaffected. This is because nurses were not involved in the laboratory process in this model.

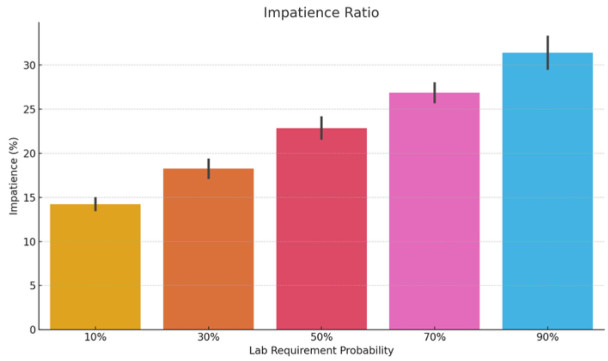
Throughput decreased steadily with higher lab loads, confirming that diagnostic delays directly hinder service capacity and patient turnover.

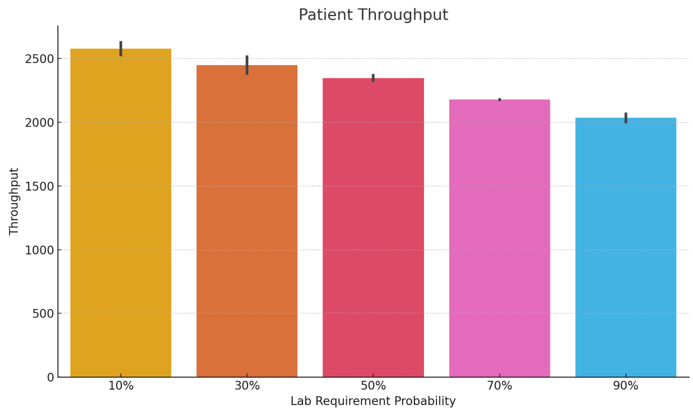
Experimental Outcomes of Laboratory Test Requirement Rate Variation:











Conclusions:

This study investigated how variations in key system parameters affect the performance of an emergency room through discrete-event simulation. The experiments focused on three core factors: doctor shift capacity, patient arrival rate, and the proportion of patients requiring laboratory tests. The findings offer valuable insights into how hospital resources and demand interact under different operating conditions.

The following key conclusions were drawn from the experiments:

- The adjustments made to the doctor's schedule had a substantial impact on reducing average triage waiting times and enhancing patient throughput. However, an increase in staffing resulted in lower doctor utilization, indicating a trade-off between capacity and efficiency.
- Higher patient arrival rates put pressure on the system, causing longer wait times, increased patient impatience, and the saturation of available resources. While throughput increased in absolute numbers, the performance per resource unit declined, highlighting capacity limitations.
- The increase in the proportion of patients requiring laboratory tests had a negative effect on almost all performance indicators. As the number of patients requiring lab work increased, delays accumulated, leading to longer total times in the system and higher abandonment rates. Doctor utilization increased, while nurse utilization remained unaffected, as nurses were not involved in diagnostic processing.

Overall, the model provided a flexible and informative simulation framework to evaluate healthcare operations. The results demonstrate the importance of balancing resource availability with demand levels to ensure timely and effective care delivery in emergency settings.