République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique
المدرسة الوطنية المتعددة التقنيات

Département du Génie Industriel

End of Study Project Dissertation
for Obtaining State Engineer's Degree in Industrial Engineering

Option : Industrial Management

Data-Driven Optimization for Nurse Scheduling and Rescheduling Problem

Hadil CHORFI

Publicly presented and defended on 30/06/2025

Composition of the jury :

President	Mr.	Ali	BOUKABOUS	MAA	ENP
Examiner	Mr.	Iskander	ZOUAGHI	MCA	ENP
Supervisor	Ms.	Samia	BELDJOUDI	MCA	ENP
Supervisor	Mr.	Yassine	OUAZENE	MCF HDR	UTT
Supervisor	Ms.	Yasmine	ALAOUCHICHE	Dr.	UTT

ENP 2025

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Algiers, Algeria.
`www.enp.edu.dz`

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique
المدرسة الوطنية المتعددة التقنيات

Département du Génie Industriel

End of Study Project Dissertation
for Obtaining State Engineer's Degree in Industrial Engineering

Option : Industrial Management

Data-Driven Optimization for Nurse Scheduling and Rescheduling Problem

Hadil CHORFI

Publicly presented and defended on 30/06/2025

Composition of the jury :

President	Mr.	Ali	BOUKABOUS	MAA	ENP
Examiner	Mr.	Iskander	ZOUAGHI	MCA	ENP
Supervisor	Ms.	Samia	BELDJOUDI	MCA	ENP
Supervisor	Mr.	Yassine	OUAZENE	MCF HDR	UTT
Supervisor	Ms.	Yasmine	ALAOUCHICHE	Dr.	UTT

ENP 2025

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Algiers, Algeria.
www.enp.edu.dz

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique
المدرسة الوطنية المتعددة التقنيات

Département du Génie Industriel

Mémoire de Projet de Fin d'Études

Pour l'obtention du diplôme d'Ingénieur d'État en Génie Industriel

Option: Management Industriel

Optimisation basée sur les données pour le problème d'ordonnancement et de réordonnancement des infirmiers

Hadil CHORFI

Présenté et soutenu publiquement le 30/06/2025

Composition du jury :

Président	M.	Ali	BOUKABOUS	MAA	ENP
Examinateur	M.	Iskander	ZOUAGHI	MCA	ENP
Encadrante	Mme.	Samia	BELDJOUDI	MCA	ENP
Encadrant	M.	Yassine	OUAZENE	MCF HDR	UTT
Encadrante	Mme.	Yasmine	ALAOUCHICHE	Dr.	UTT

ENP 2025

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Alger, Algérie.
www.enp.edu.dz

الملخص

تُعد جدولة الممرضين في المستشفيات مهمةً شديدة التعقيد، نظراً لارتباطها بقيود تنظيمية وتعاقدية وتشغيلية صارمة، إلى جانب مستويات عالية من عدم اليقين. فرغم أن النماذج التقليدية في البرمجة الرياضية قادرة على إنتاج جداول عمل أولية قابلة للتطبيق، إلا أنها غالباً ما تعجز عن التعامل مع الاضطرابات غير المتوقعة، مثل الغيابات المفاجئة في اللحظات الأخيرة. وتؤدي هذه الغيابات إلى تدهور جودة الرعاية الصحية، وخلل في توزيع عبء العمل، مما يستدعي اتخاذ قرارات تصحيحية عاجلة ومكلفة. كما أن معظم النماذج الحالية نادراً ما تتضمن آليات تنبؤية أو تدابير استباقية لمعالجة هذا النوع من التقلبات. ويكمن التحدي الرئيسي الذي يتناوله هذا العمل في تصميم نظام ذكي للجدولة وإعادة الجدولة، قادر على استباق الغيابات اليومية والتعامل معها بمرونة وفعالية، مع الحفاظ على عدالة توزيع العمل، والامتثال للقيود التنظيمية، وضمان جودة التغطية التمريضية..

الكلمات المفتاحية: جدولة الممرضين، إعادة الجدولة، خوارزمية تقريبية، الغيابات، نموذج هيردل، النماذج التنبؤية، البرمجة متعددة الأهداف.

Résumé

La planification des horaires des infirmiers en milieu hospitalier est une tâche complexe, soumise à de nombreuses contraintes réglementaires, contractuelles et opérationnelles. Les modèles classiques d'optimisation permettent d'élaborer un planning initial faisable, mais ils restent inadaptés face aux absences imprévues de dernière minute. Ces perturbations dégradent la qualité des soins, déséquilibrent la charge de travail et nécessitent des ajustements coûteux en urgence. Très peu d'approches actuelles intègrent des mécanismes prédictifs ou proactifs pour y faire face. La problématique centrale de ce mémoire consiste donc à concevoir un système de planification et de replanification capable d'absorber les absences quotidiennes de manière réactive et robuste, tout en minimisant les perturbations du planning initial.

Mots-Clés : Planification des infirmiers, replanification, heuristique, absence, hurdle, modèle prédictif, programmation multi-objective.

Abstract

Nurse scheduling in hospitals is a highly constrained and uncertain task. While traditional optimization models can generate feasible baseline schedules, they often fail to account for unplanned disruptions such as last-minute absences. These absences compromise care quality, create workload imbalances, and force costly last-minute adjustments. Existing models rarely integrate predictive insights or proactive mechanisms to handle such volatility. The core challenge addressed in this work is to design a scheduling and rescheduling system that anticipates and reacts to daily absences with minimal disruption, while maintaining fairness, regulatory compliance, and staffing quality.

Keywords: Nurse scheduling, rescheduling, heuristic, absence, hurdle model, predictive modeling, multi-objective programming.

Acknowledgements

I would like to sincerely thank my academic supervisors: Mr. Yassine Ouazene, Ms. Yasmine Alaouchiche, and Ms. Samia Beldjoudi, for their valuable support and guidance throughout this internship. I am particularly grateful to Mr. Ouazene for his availability, clear explanations, and constructive advice. His support helped me better understand many technical aspects, and his structured approach made our online meetings efficient and productive.

I also wish to thank Ms. Alaouchiche for her relevant suggestions and her help in solving several difficulties, especially during the early phases of the project. Her guidance clearly contributed to improving the direction and quality of this work.

My appreciation goes as well to Ms. Beldjoudi for her continuous support and for encouraging me to pursue this research opportunity. Her involvement was very helpful throughout the internship.

I would like to extend my thanks to all the professors who contributed to my academic journey. I am especially thankful to Mr. Zouaghi for his constant availability and dedication. His support went beyond teaching, and he was always ready to assist students whenever needed.

Finally, I would like to acknowledge Mr. Boukabous, Ms. Bouchafaa, Ms. Bareche, Mr. Gourine, Ms. Bouazza, and Mr. Fourar for their teaching and their support during my studies. Their contribution played an important role in my academic and personal development.

Hadil CHORFI

Dedication

First and foremost, I am deeply grateful to God, whose guidance and mercy have accompanied me through every step of my life.

I also want to thank myself, for rising after every fall, pushing through every moment of doubt, and becoming stronger, more determined, and more resilient each time.

To my mother, the most precious person in my life, your sacrifices, love, and unwavering support shaped the person I am today. You are the foundation of all my success.

To my father, your love and sacrifice are equal in every way. You work tirelessly every day to give me everything I need. I promise: you will always be proud of me.

To my grandfather, who left us too soon. I have finally become the engineer you always hoped to see one day. I will never forget the values and lessons you taught me. You remain alive in my heart, always.

To my little sister, even though you're younger, you've done more for me than you know. Thank you for being there.

To my little brother, my source of comfort, my listener, my cheerleader. I can't imagine this journey without you.

To my extended family, even from afar, your presence and affection were always felt. Thank you for being there in your own way.

To Asma, my soul sister since childhood. Thank you for all the hours of listening, for your care, and for standing by me. Your friendship is a blessing.

To Samia, the most precious gift of my ENP journey. Your genuine support and presence this year in every step meant the world to me.

To Serine, your strength has always been a source of motivation for me. Thank you for being there, truly.

To Lyna, even though our friendship is recent, you've shown a closeness and support that feel like they've always been there. I'm truly grateful for you.

To all those with whom I shared unforgettable moments at ENP, my second home: Mehdi, whose sense of humor made difficult days a little lighter. Samir, the kindest friend. Idir, the most loved one, thank you for everything, especially your help at the very beginning of this project. Sofiane, for your wise advice. To the kindest people ENP brought into my life, I will always carry the warmth of your friendship in my heart.

And finally, to everyone not mentioned by name but who holds a place in my heart, your love and support meant more than words can say.

Hadil CHORFI

Contents

List of Tables

List of Figures

List of Acronyms

- **CPLEX** : IBM Optimization Solver for Linear and Integer Programming
- **EDA** : Exploratory Data Analysis
- **KPI** : Key Performance Indicator
- **LR** : Linear Regression
- **MAE** : Mean Absolute Error
- **MILP** : Mixed Integer Linear Programming
- **NSGA-II** : Non-dominated Sorting Genetic Algorithm II
- **NSP** : Nurse Scheduling Problem
- **OR** : Operations Research
- **RF** : Random Forest
- **RMSE** : Root Mean Squared Error
- **XGB** : Extreme Gradient Boosting

General Introduction

Over the past decades, healthcare systems worldwide have faced growing pressure to optimize resources while maintaining high standards of patient care. Among the most critical and resource-intensive components of hospital operations is the management of the nursing workforce. Nurses represent not only the largest group of healthcare professionals but also the most operationally constrained due to labor laws, contractual regulations, and the continuous nature of care. As such, **nurse scheduling** has long been recognized as a key logistical and strategic challenge in hospital management.

Initially approached using manual methods and administrative experience, the field of nurse scheduling evolved significantly in the 1990s and 2000s with the advent of mathematical optimization techniques. **Integer programming**, **constraint programming**, and later **metaheuristics** became the dominant tools for generating fair, feasible, and efficient schedules. These models aimed to balance supply (nurses' availability and preferences) with demand (patient care needs) while adhering to institutional rules. Many successful systems were deployed to generate optimal or near-optimal *baseline schedules*, planned weekly or monthly.

However, these systems were largely static, assuming that once the schedule is set, it can be followed as-is. In real-world hospitals, this is rarely the case. Unexpected events such as sudden absences, patient influx, or shift swaps routinely disrupt the planned roster. The rise of pandemics (e.g., COVID-19), seasonal fluctuations, and staffing shortages has only made such disruptions more frequent and harder to manage. These unforeseen events have exposed a critical gap in traditional scheduling systems: their inability to adapt dynamically.

To address this challenge, the field has progressively shifted towards **rescheduling**, the ability to update schedules in real time to reflect new constraints and realities. This transition requires moving from static optimization to **dynamic, responsive systems** that can proactively plan for uncertainty and react swiftly when disruptions occur. It also calls for an integration of **predictive analytics**, drawing on historical data to identify high-risk situations and prepare mitigation strategies in advance.

In this era of **data-driven decision-making** and **intelligent automation**, new approaches are being developed that go beyond the initial schedule to embed **resilience** and **flexibility** into the planning process. These include the use of predictive models to anticipate absenteeism, hybrid frameworks that combine optimization with machine learning, and real-time heuristics to adjust assignments as the situation evolves.

Beyond their theoretical contributions, these intelligent rescheduling solutions hold concrete operational value for hospitals. By allowing timely and intelligent adjustments to sudden absences, these solutions limit the reliance on rushed last-minute fixes, which can lead to poor shift balance, unfair workloads, and costly overtime. Proactively protecting critical shifts with buffer staff based on historical risk patterns allows better resource allocation and lowers the risk of under-coverage, which directly impacts patient care quality. Furthermore, the ability to simulate various disruption scenarios such as epidemic surges or chronic absenteeism offers hospital managers a strategic tool to plan contingency responses and validate staffing policies. Ultimately, these solutions contribute to creating more sustainable and resilient healthcare systems, better equipped to face operational uncertainty.

It is in this evolving context that our final-year project is situated, addressing the following guiding question:

"How can we design an intelligent nurse scheduling and rescheduling system that effectively absorbs staff absences with minimal disruption to the original plan?"

To answer this, we propose a **two-stage hybrid approach**:

- A **baseline optimization model**, built using Mixed-Integer Linear Programming (MILP), which integrates a *criticality score* computed from historical absence data. This score identifies shifts with high absenteeism risk, where overstaffing is introduced proactively as a buffer.

- A **lightweight rescheduling heuristic**, executed in real time, that reallocates surplus nurses (initially assigned to a virtual overflow unit) to fill uncovered shifts without recalculating the full schedule.

The proposed solution is embedded in a **rolling horizon simulation framework**, where each day:

1. Staff absences are either drawn from historical records or generated via a predictive Hurdle-based model.
2. The heuristic reassigns nurses from the overflow unit to cover these absences.

This simulation enables robust evaluation of the system's performance across different absence scenarios, normal and crisis conditions and facilitates comparison between our data-driven solution and a conventional baseline.

The structure of this work is as follows:

- **Chapter 1: State of the Art**: Presents a comprehensive review of the literature on nurse scheduling and rescheduling, from classical optimization to data-driven and hybrid approaches. We position our contribution relative to recent works that emphasize adaptability and operational robustness.
- **Chapter 2: Optimization Model and Algorithmic Framework**: Describes the formulation of our MILP model, the design of the criticality score, and the rescheduling heuristic using an overflow unit, all within a modular and flexible simulation environment.
- **Chapter 3: Computational Study**: Evaluates the effectiveness of the proposed solution using real hospital data under various absence conditions. We analyze metrics such as absence coverage, preference satisfaction, under- and overstaffing, and planning stability.

We conclude with a discussion of the implications of our work, its relevance to modern hospital operations, and perspectives for future research on intelligent, resilient workforce planning in healthcare.

Chapter 1

State of the Art

Introduction

This chapter provides the theoretical and methodological foundations for our work on nurse planning in dynamic healthcare environments. We begin with fundamental concepts from operations research and scheduling theory, which support the development of optimization models.

The focus then narrows to healthcare-specific planning problems, notably the Nurse Scheduling Problem (NSP), which must account for institutional rules, staff preferences, and operational uncertainty. We review the main modeling approaches including deterministic, stochastic, and hybrid methods and examine how recent contributions address unpredictability through proactive and reactive mechanisms.

A dedicated section explores the need for rescheduling, which is increasingly relevant in the face of unexpected absences or demand fluctuations. Finally, a review of the literature highlights representative works and positions our contribution: a data-driven two-stage framework aimed at improving staffing resilience with minimal disruption.

1.1 Background Theory

1.1.1 Operations Research

Operations research (OR) is a multidisciplinary field that applies mathematical and analytical techniques to improve decision-making. Emerging during World War II, OR has evolved to address complex problems across many sectors. The field includes various techniques such as simulation, optimization, linear programming, game theory, and search theory. OR's main goal is to find optimal or near-optimal solutions using these methods.

OR has wide-ranging applications in global health, including health systems, clinical medicine, public health, and health innovation .

Mathematical Programming

Mathematical programming is a branch of operations research that helps optimize resource allocation. It involves constructing a mathematical model of a real-world problem and determining the optimal solution, typically by minimizing a cost or maximizing a benefit, while satisfying certain constraints.

Let:

- $x \in \mathbb{R}^n$ be the vector of decision variables,
- $f(x)$ be the objective function,
- $g_i(x) \leq 0$, for $i = 1, ..., m$, be the inequality constraints,
- $h_j(x) = 0$, for $j = 1, ..., p$, be the equality constraints.

Then, the general form of a constrained optimization problem is:

$$
\begin{aligned}
\text{Minimize (or Maximize)} \quad & f(x) \\
\text{subject to} \quad & g_i(x) \leq 0 \quad \text{for } i = 1, ..., m \\
& h_j(x) = 0 \quad \text{for } j = 1, ..., p \\
& x \in X
\end{aligned}
$$

Where:

- $f(x)$ represents the objective to optimize,
- $g_i(x)$, $h_j(x)$ represent the constraints on the variables,
- $X \subseteq \mathbb{R}^n$ is the feasible set, including domain restrictions like non-negativity or bounds on variables.

Multi-Objective Optimization and Modeling

Among the various problem types, *multi-objective optimization* involves the simultaneous optimization of several potentially conflicting objectives. In such cases, the participation of a decision maker is often essential to identify acceptable trade-offs [2].

Solution approaches for multi-objective problems are generally categorized based on when the decision-maker's preferences are incorporated:

- **A priori**, where preferences are set before solving the problem,
- **Interactive**, where preferences are refined during the solution process,
- **A posteriori**, where a set of solutions is first generated, and preferences are applied afterward [3].

The lexicographic method, for example, represents an a priori approach, in which objectives are prioritized and solved sequentially.

In terms of modeling, mathematical programming models can be either:

- **Deterministic**, assuming all parameters are known with certainty,
- **Stochastic**, accounting for uncertainty using probability distributions [4].

For complex stochastic systems where the objective functions cannot be expressed analytically, simulation optimization techniques offer a powerful alternative. These methods often incorporate interactive algorithms that facilitate the resolution of multi-objective problems under uncertainty [5].

1.1.2 Scheduling Theory

Scheduling problems constitute a fundamental domain within operations research and artificial intelligence concerning the temporal allocation of resources to tasks subject to various constraints. One common method for addressing scheduling problems is through constraint satisfaction techniques. These techniques aim to find solutions that satisfy a set of constraints, such as precedence relations, resource limits, or timing requirements. This approach is particularly useful due to its flexibility in representing complex problem structures [6].

Static vs. Dynamic Scheduling

Traditional scheduling approaches are often static. In static scheduling, a schedule is generated in advance and is assumed to remain fixed throughout execution. However, in real-world environments, unexpected disruptions, delays, or changes in resource availability often occur. This makes static approaches less effective for practical applications [7].

Dynamic scheduling has emerged as an alternative. Continually updates schedules based on real-time information, allowing systems to adapt to unforeseen events and improve resilience and performance [7].

Techniques for Dynamic Scheduling

To support dynamic scheduling, various techniques have been proposed. In the following, we present the most prominent ones.

Heuristics algorithms are simplified, rule-based approaches that aim to find good-quality solutions within a reasonable computational time. They are particularly useful for complex problems where exact methods are too slow or computationally infeasible.

Many real-world scheduling problems are combinatorial in nature, resulting in large and complex search spaces. As highlighted by Juan et al. [9], solving such problems using exact optimization techniques can be impractical due to the excessive time and resources required. In these cases, heuristics provide a more efficient way to obtain feasible and near-optimal solutions.

Heuristics often rely on intuitive strategies and prior knowledge to guide the search process. Their design can range from simple decision-making rules to more advanced iterative procedures that progressively construct a solution [8, 11].

A well-known category of heuristics is greedy algorithms, which iteratively select the most beneficial component at each step based on a specific evaluation criterion [10]. Although they offer fast and easy implementations, they do not guarantee optimal results and may converge to suboptimal solutions.

Despite these limitations, heuristics are widely adopted due to their practicality and ability to deliver satisfactory results within a limited time frame.

Meta-heuristics such as genetic algorithms or simulated annealing are designed to explore large and complex solution spaces efficiently. Unlike basic heuristics, meta-heuristics incorporate mechanisms to avoid local optima and improve the chances of finding near-global solutions.

Multi-agent Systems involve several intelligent agents that collaborate or compete to build, evaluate, and adjust schedules dynamically. These systems are particularly relevant in environments requiring decentralized decision-making and high adaptability.

Artificial Intelligence Methods AI-based techniques, including machine learning and adaptive algorithms, aim to improve scheduling performance over time by learning from historical data and evolving through experience [7]. These methods are gaining popularity in highly dynamic or uncertain scheduling environments.

1.1.3 Healthcare Scheduling

Healthcare scheduling is a complex field that covers various aspects of patient care coordination. Recent research has focused on capacity planning in hospital units, outpatient clinics, and healthcare networks, as well as appointment scheduling, surgery, and workforce management [12]. These scheduling problems are challenging due to uncertainty in healthcare environments, including unpredictable patient arrivals, emergency cases, variable treatment times, and staff availability.

Effective healthcare scheduling aims to balance resource utilization, patient satisfaction, and clinical outcomes while working within operational and regulatory constraints.

Types of Healthcare Scheduling Problems

Patient Admission Scheduling (PAS) : is a critical subdomain of healthcare scheduling that involves determining the optimal assignment of patients to hospital beds or units, often across multiple departments. This process must respect various clinical, operational and administrative constraints, such as bed availability, patient urgency, medical specialties, infection control protocols, and patient preferences. The complexity of PAS increases due to the dynamic and uncertain nature of hospital environments, where emergency admissions, discharges, and changes in patient condition can disrupt planned schedules.

Unlike basic bed allocation, PAS is typically performed in advance and must account for both elective and urgent patients, requiring a balance between operational efficiency and quality patient care. It plays a key role in managing hospital flow, reducing overcrowding, avoiding last-minute cancellations, and ensuring timely treatment. Properly executed admission scheduling can improve continuity of care and resource utilization across different hospital units, including intensive care, surgery, and general medicine.

Recent research in PAS reflects its significance in modern healthcare management. Studies such as those by Liu et al. [13], Demeester et al. [14], Bamigbola et al. [15], and Abdalkareem et al. [16] have explored increasingly sophisticated methods to enhance the decision-making process in PAS.

Nurse Scheduling: In healthcare, efficient staff allocation is essential to maintain continuous high-quality care. The Nurse Scheduling Problem (NSP) involves assigning shifts to nurses over a planning horizon while respecting institutional rules, contractual constraints, and individual preferences [17].

What makes NSP particularly challenging is its multidimensional nature: schedules must ensure coverage across all shifts and units, account for varying qualifications, comply with labor regulations, and consider staff fairness. Combined with the 24/7 nature of hospital operations and unpredictable disruptions such as absences, this results in a highly constrained, combinatorial problem.

The next sections delve into the types of constraints, terminologies, and optimization strategies that have emerged to address this complex challenge.

Operating Room Scheduling : Another critical category of healthcare scheduling is the management of operating rooms (ORs), which involves coordinating elective and emergency surgical procedures. OR scheduling is particularly complex due to the need to balance planned surgeries with the unpredictable nature of emergency cases. This challenge is especially evident in Level-1 trauma centers, where urgent interventions are frequent and time-sensitive [18].

One of the key difficulties in the scheduling of the operating room is managing uncertainty, such as variability in the duration of surgery, last-minute changes in patient conditions, and the limited availability of medical personnel and equipment. Effective scheduling requires close coordination among surgical teams, anesthesiologists, nurses, and specialized resources, while ensuring patient safety and continuity of care.

Beyond operational concerns, OR scheduling has a significant impact on broader hospital performance metrics. It influences patient wait times, the risk of surgery cancellations, the distribution of staff workload, and the overall utilization of costly resources such as operating theaters.

Recent research emphasizes the importance of integrated planning that accounts for the interdependencies between operating rooms, healthcare professionals, and surgical procedures. For example, Tsang et al. [19] highlight the relevance of coordinated scheduling frameworks that are robust to uncertainty and adaptable to real-time changes in the surgical environment.

Other Scheduling Problems

In addition to the core categories of healthcare scheduling, several other types also play an essential role in ensuring efficient and responsive care delivery.

Physician scheduling involves assigning doctors shifts or on-call duties between departments, often considering specializations, availability, work-hour regulations and preferences. This type of scheduling must ensure adequate coverage, especially in departments with limited staff or a high volume of patients.

Home healthcare scheduling focuses on planning visits by caregivers or nurses to patients' homes. It must consider travel time, visit duration, patient needs, staff skills, and geographical constraints, making it similar to vehicle routing problems with added healthcare-specific considerations.

Telemedicine scheduling has emerged more recently with the rise of digital health services. It involves coordinating virtual appointments between patients and healthcare providers, often between different time zones or platforms. While more flexible in terms of logistics, it still requires attention to provider availability, consultation time slots, and technology readiness.

1.1.4 Data Analysis

Data analysis is a multifaceted process of building understanding and extracting insights from complex datasets, often similar to cognitive sensemaking [48]. In the context of data-driven scheduling, this involves exploring and interpreting operational or workforce-related data potentially structured as complex or functional objects [49] to inform and optimize scheduling decisions. Techniques such as exploratory data analysis (EDA) are especially valuable in identifying hidden patterns, trends, and constraints that affect scheduling outcomes [50], while addressing issues such as incomplete information and bias [51]. This highlights the importance of flexible, context-aware methods and interdisciplinary collaboration in modern data-driven scheduling practices.

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a foundational approach in statistics, aimed at examining datasets to uncover underlying structures, detect anomalies, and formulate hypotheses without relying on prior assumptions. First introduced by John W. Tukey, EDA emphasizes graphical techniques such as histograms, boxplots, and scatterplots to visualize data distribution and variability [56, 57]. It is typically applied before confirmatory analysis and is considered essential for guiding model development, especially when dealing with complex or high-dimensional datasets [59]. Beyond visualization, EDA also includes methods such as Exploratory Factor Analysis (EFA), which are useful for identifying latent constructs and reducing dimensionality [58]. Despite the advent of automated analytics, EDA remains a critical first step in any rigorous data science or machine learning pipeline.

Predictive Data Analysis

Predictive analytics involves the use of statistical models and machine learning techniques to forecast future events based on historical data [60, 61]. In healthcare workforce planning, such methods are particularly useful for anticipating staff absences, a key source of operational uncertainty. By identifying recurring patterns and risk factors, predictive models enable proactive scheduling strategies that can reduce last-minute disruptions. While predictive analytics has been widely applied in public health and operational planning [62], its integration into critical systems such as hospital staffing requires both methodological rigor and interpretability to ensure safe and effective decision-making [60].

Among the many predictive models applicable to count data, one frequently encountered challenge is dealing with excess zeros common in operational healthcare datasets such as absenteeism records. To address this, specialized models such as the hurdle model have been developed.

Count-Based Modeling with the Hurdle Model

When modeling count data, such as daily staff absences, a common issue is the presence of excess zeros and asymmetric distributions. Standard models like Poisson or Negative Binomial regressions often fail to capture this structure, particularly when a large proportion of observations correspond to structural zeros. The hurdle model addresses this limitation by splitting the process into two components: one models the probability that an event occurs, and the other estimates the count, conditional on its occurrence. This approach has shown excellent performance in fields such as dental epidemiology [63], mental health [64], HIV treatment analysis [65], and financial risk modeling [66].

The mathematical formulation of the hurdle model reflects its two-part nature. Let us define the general form of the model and its most common variants.

Formally, let Y_i denote the count response for observation $i = 1, \ldots, n$. The general structure of a hurdle model is the following:

$$P(Y_i = y_i) = \begin{cases} p_i & \text{if } y_i = 0, \\ (1 - p_i)\dfrac{p(y_i; \mu_i)}{1 - p(0; \mu_i)} & \text{if } y_i > 0, \end{cases} \tag{1.1}$$

where p_i is the probability of a structural zero, $p(y_i; \mu_i)$ is the probability mass function (PMF) of a count distribution with mean μ_i, and $p(0; \mu_i)$ is the PMF evaluated at zero.

If the count component follows a Poisson distribution, the hurdle Poisson model becomes:

$$P(Y_i = y_i) = \begin{cases} p_i & \text{if } y_i = 0, \\ (1 - p_i)\dfrac{e^{-\mu_i}\mu_i^{y_i}/y_i!}{1 - e^{-\mu_i}} & \text{if } y_i > 0. \end{cases} \tag{1.2}$$

To account for overdispersion, a Negative Binomial (NB) distribution is often used. The hurdle NB (HNB) model is given by:

$$P(Y_i = y_i) = \begin{cases} p_i & \text{if } y_i = 0, \\ \dfrac{1 - p_i}{1 - \left(\dfrac{r}{\mu_i + r}\right)^r} \cdot \dfrac{\Gamma(y_i + r)}{\Gamma(r) y_i!} \left(\dfrac{\mu_i}{\mu_i + r}\right)^{y_i} \left(\dfrac{r}{\mu_i + r}\right)^r & \text{if } y_i > 0, \end{cases} \tag{1.3}$$

where r is the dispersion parameter, and $\Gamma(\cdot)$ denotes the gamma function.

Covariates can influence both the zero-generation and count processes through two link functions:

$$\log(\mu_i) = \mathbf{x}_i^\top \alpha, \quad \operatorname{logit}(p_i) = \mathbf{z}_i^\top \beta, \tag{1.4}$$

where α and β are the regression coefficients for covariates $\mathbf{x}_i$ and $\mathbf{z}_i$ respectively [67].

The Interaction Between Data Analytics and Operations Research

The integration of data analytics and operations research (OR) is transforming decision-making in complex domains such as healthcare. OR methods such as linear programming, Markov Decision Processes, and simulation are widely applied to optimize various aspects of healthcare delivery, including diagnosis, treatment, organ transplantation, and patient flow [52]. These approaches help address challenges related to hospital resource organization, surgical scheduling, and healthcare facility location planning [53]. The growing availability of electronic health records and other observational data has enabled the development and validation of data-driven OR models [54]. Recent research also highlights the importance of optimization in advancing both predictive and prescriptive analytics, such as through optimal estimation of clustered models and the optimization of objective functions derived from tree ensemble models [55]. This interaction between data analytics and OR is essential for creating sustainable, long-term solutions in healthcare management and policy.

1.2 Literature Review

1.2.1 Key Terminology in Nurse scheduling

Before diving into the different approaches found in the literature, it is helpful to clarify some commonly used terms in the field of nurse scheduling. These concepts serve as a foundation for understanding how scheduling models are built and evaluated [20].

Planning Period: This refers to the total time span covered by the schedule. It can vary widely from a few weeks to several months and defines the window within which all shifts must be assigned.

Skill Category: Nurses are often grouped according to their qualifications, certifications, or responsibilities, for example, general nurses, specialized nurses, or team leaders. These categories must be taken into account to ensure the right mix of skills is present in each shift.

Shift Type: Shifts are usually divided by start and end times, such as morning (e.g. 7:00 am – 3:00 pm), evening (e.g. 3:00 pm – 10:00 pm), or night (e.g. 10:00 pm – 7:00 am). Each shift type has its own staffing and regulatory considerations.

Work Regulations: These are the contractual or institutional rules that define how much and how often a nurse can work. For instance, one nurse might work five days a week, while another is scheduled for six, depending on their contract.

Hard Constraints: These are rules that must always be respected in any valid schedule. Examples include minimum staffing levels, mandatory rest periods, or restrictions on maximum working hours.

Soft Constraints: These represent preferences or flexible guidelines such as shift requests or fairness in shift distribution that can occasionally be violated, but typically with a penalty. Balancing these constraints helps improve staff satisfaction and schedule quality.

Coverage: This refers to the number of nurses needed for each shift and skill category. Ensuring proper coverage is essential for maintaining safe and effective patient care.

Time Restrictions: These are limitations designed to prevent overwork and promote fair distribution of shifts, such as restrictions on consecutive workdays or ensuring adequate rest between shifts.

Request: Nurses can express preferences regarding their schedules, for instance, asking for a specific day off or requesting not to work nights if possible.

1.2.2 Problem Constraints in Nurse scheduling

In nurse scheduling, constraints are generally divided into two main categories: *hard constraints* and *soft constraints*. Hard constraints are mandatory rules that must always be respected if even one of them is violated, the resulting schedule is considered invalid. Soft constraints, on the other hand, are more flexible. They can be violated, but doing so usually comes at a cost that is measured by a penalty function. The goal is to generate schedules with the lowest possible penalty, meaning a better balance between feasibility and staff preferences [21, 22].

Typically, coverage requirements such as ensuring that each shift has the right number of nurses with the appropriate skills are modeled as hard constraints. Meanwhile, many time-related constraints (such as preferred rest days or avoiding long stretches of work) are treated as soft constraints [23].

Soft constraints are often grouped into three broad types:

Series Constraints: These limit how often something can occur in a row, for example, a nurse working too many consecutive night shifts or having too many consecutive days off.

Successive Series Constraints: These manage transitions between patterns, like how many workdays can be followed by rest days, or vice versa.

Counter Constraints: These set limits on how many times a specific event can occur over a certain period, such as the total number of working hours, weekend shifts, or off-days over a planning horizon.

As noted by Burke and Curtois [33], the specific types of constraints applied often depend on hospital regulations and practical considerations. However, some constraints are commonly found across the literature and scheduling models [20], as shown in Table 1.1:

Code	Constraint Description
C1	Minimum/maximum workload
C2	Minimum/maximum consecutive working days
C3	Minimum/maximum consecutive days off
C4	Minimum/maximum/exact number of identical consecutive shifts
C5	Minimum/maximum number of shifts during weekends
C6	Minimum/maximum number of consecutive working weekends
C7	Minimum/maximum number of shift rotations
C8	Minimum/maximum rest time between two shifts
C9	Minimum/maximum/exact number of working hours (e.g., 6–8 hours/day)
C10	Nurse skill or qualification categories
C11	Individual nurse-specific requirements and preferences
C12	Scheduling history (e.g., previously assigned shifts)
C13	Constraints related to specific shift types
C14	Nurse requests for specific days off/on
C15	Nurse requests for specific shifts off/on
C16	Leave or holidays (e.g., annual leave, sick leave)
C17	Group constraints (e.g., nurses who must or must not work together)
C18	Shift-related constraints (e.g., no double shifts, no overlapping)

Table 1.1: Common nurse scheduling constraints (adapted from Ngoe, 2022)

1.2.3 Optimization Approaches for Nurse Scheduling

Nurse scheduling is a complex problem influenced by multiple constraints, ranging from institutional policies to staff preferences and unpredictable events. To address this complexity, a wide variety of optimization approaches have been developed in the literature. These approaches can be broadly categorized into deterministic, stochastic, and hybrid methods, each offering distinct advantages depending on the nature of the scheduling environment and the objectives pursued.

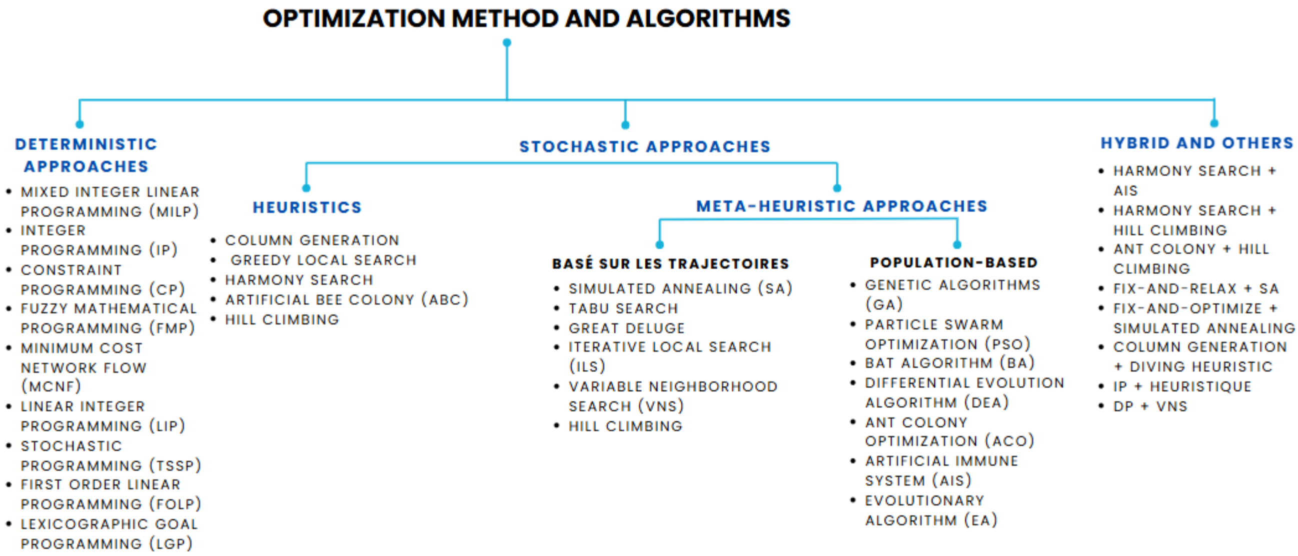

Figure 1.1: Nurse Restoring solutions method

Deterministic Approaches

Deterministic approaches in optimization refer to methods in which all the data involved in the problem (such as costs, resources, and time) are known with certainty and do not change. In these approaches, the model behavior is fully predictable and the same set of inputs will always lead to the same result. Mathematical programming is a widely used deterministic approach with two main types: Linear Programming (LP) and Nonlinear Programming (NLP). LP involves only linear functions and is widely used in problems where the relationships between variables are straightforward. On the other hand, NLP deals with more complex relationships and can accommodate both discrete and continuous variables. These deterministic approaches assume that all parameters are fixed and known, which simplifies the problem-solving process, ensuring that the solutions obtained are repeatable and reliable [20].Here are some of the deterministic approaches commonly used in nurse scheduling.

INTEGER PROGRAMMING

Integer Programming (IP) is a specific form of LP in which all decision variables are constrained to take integer values. It is often applied in nurse scheduling due to its ability to precisely model scheduling constraints. Zanda et al. proposed a linear integer programming approach for long-term nurse scheduling in the surgery department of a university hospital in Italy. Their model handled sudden changes in nurse availability and allowed real-time schedule updates through a Decision Support System [25].

Branch-and-price

Branch-and-price algorithms have been successfully applied to various nurse scheduling problems. These approaches typically model the problem as an integer program and use column generation to solve large instances efficiently [26, 27]. The pricing subproblem often involves solving a shortest path problem with resource constraints, where dominance rules and acceleration techniques can significantly improve performance [27]. Some models use rotations or sequences of consecutive worked days as columns, which can be effective for certain problem variants [26]. Others employ context-free grammars to model complex shift construction rules [28]. These approaches have demonstrated success in solving various nurse scheduling problems, including those with multiple activities, tasks, and personalized constraints [28, 27].

Constraint Programming

Constraint Programming (CP) has emerged as a powerful approach for addressing complex scheduling challenges such as nurse scheduling. Unlike traditional optimization techniques, CP emphasizes the satisfaction of constraints over the optimization of objective functions, making it particularly suitable for generating feasible schedules under intricate rules and regulations [29]. One notable application is the introduction of a bound-consistent spread constraint, which allows an even distribution of values across variable vectors, improving the ability of the model to balance the workload and respect institutional policies [29].

Lexicographic Goal Programming

Lexicographic Goal Programming (LGP) has been increasingly utilized in nurse scheduling problems as a powerful multi-objective optimization technique. It allows decision-makers to prioritize conflicting objectives by assigning a hierarchical order of importance to each goal. In the context of healthcare, LGP has been particularly effective in balancing operational efficiency with staff and patient satisfaction. For example, Lim et al. [30] applied LGP to minimize labor costs, patient dissatisfaction, and nurse idle time, while simultaneously maximizing job satisfaction. This prioritization framework ensures that more critical goals, such as maintaining adequate nurse-patient ratios, are addressed before less urgent concerns like individual preferences. Furthermore, Ang et al. [31] demonstrated the applicability of LGP in emergency departments, where the complexity of scheduling increases due to fluctuating demand and high pressure environments. Their study focused on optimizing shift preferences and the equitable distribution of rest days, showing that LGP can effectively manage trade-offs between organizational constraints and personal preferences. By structuring goals lexicographically, healthcare institutions can generate feasible and fair schedules that reflect both institutional priorities and human factors.

Stochastic approaches in nurse scheduling address the complexities of resource allocation and the inherent uncertainty in healthcare environments. These methods consider fluctuations in patient demand, staff availability, and unforeseen events such as last-minute absences or surges in admissions. The core objective of stochastic nurse scheduling models is to optimize staffing costs, minimize overtime, and ensure a high level of patient care quality. Such models often explicitly integrate uncertainty through probability distributions or scenario-based analyses, making them more realistic than deterministic models in dynamic healthcare contexts.

Stochastic Approaches

Stochastic approaches can be classified into two main categories: heuristic and metaheuristic methods. Heuristics provide problem-specific strategies that are computationally efficient and offer quick solutions, although they may not guarantee optimality. Metaheuristics, on the other hand, are higher-level procedures designed to explore the solution space more extensively and avoid local optima, making them suitable for complex and large-scale scheduling problems.

For instance, hybrid evolutionary algorithms, which combine genetic algorithms with techniques like simulated annealing, have demonstrated significant potential in solving nurse scheduling problems [32]. These metaheuristic frameworks exploit the strengths of each component: the global search capability of genetic algorithms and the local refinement ability of simulated annealing, enabling the discovery of high-quality schedules even under uncertainty.

Moreover, stochastic programming, particularly two-stage stochastic models, allows decision-makers to plan under uncertainty by considering both initial staffing decisions and subsequent recourse actions. Aydas et al. [34] demonstrated that such models could dynamically adjust short-term nurse schedules to better align with fluctuating patient demand. Their results indicated that this approach could reduce total staffing costs by up to 18%, highlighting the practical benefits of incorporating uncertainty into scheduling frameworks.

Hybrid Approaches

Hybrid approaches in nurse scheduling combine traditional methods with advanced optimization techniques to take advantage of their complementary strengths. A method is considered hybrid if it integrates two or more techniques, either within the same step or across different stages, especially when combining methods from different categories, such as metaheuristics and hyper-heuristics [35, 39].

For instance, Burke et al. [35] developed a hybrid method called Time Predefined Variable Depth Search (VDS), which improved initial solutions created using a greedy algorithm. This approach performed better than other hybrid methods in most test cases. Awadallah et al. [36] proposed a Hybrid Artificial Bee Colony (HABC) algorithm by combining ABC and Hill-Climbing Optimization (HCO), achieving 37 new best-known results on the INRC-I benchmark.

Jin et al. [37] studied two hybrid methods that combine the Harmony Search Algorithm (HSA) with Artificial Immune Systems (AIS). Their results showed that the cooperation-based hybrid (CHSAIS) produced better average results, although it required more computation time. Another example is the Hybrid Harmony Search Algorithm (HHSA), which integrates HSA, HCO, and Particle Swarm Optimization (PSO), achieving 38 best-known results out of 69 instances [39].

Finally, Chen and Zeng [38] applied a hybrid approach combining decision trees, greedy search, the Bat Algorithm (BA), and PSO to a real hospital scheduling problem. Their method improved both the quality of solutions and the computation speed by using decision trees to generate good initial solutions and metaheuristics for further improvement.

1.2.4 Uncertainty in Nurse Scheduling

Uncertainty is one of the main problems in nurse scheduling. It is caused by factors such as unexpected changes in patient demand, sudden staff absences, and variations in hospital service needs [47]. These factors make it difficult to create fixed schedules that work well in all situations. Traditional deterministic models assume that everything is known in advance, but in real healthcare settings, this is rarely true. As a result, these models often create schedules that are not flexible and cannot respond well to changes. This leads to the need for last-minute fixes, which can increase costs, reduce efficiency, and put extra stress on both managers and nurses. To deal with uncertainty, more advanced methods like stochastic models, robust optimization, or simulation-based planning are used. These approaches try to create schedules that are more flexible and can adapt to changes, reducing the impact of unexpected events, and improving the overall stability of staffing plans.

1.2.5 Nurse Rescheduling under Uncertainty

Nurse rescheduling has become an essential extension of traditional scheduling due to the unpredictable nature of hospital operations. Unplanned absences caused by illness, personal leave, or emergencies can disrupt initial rosters and cause imbalances in workload and coverage. To address this challenge, researchers have proposed various reactive and preventive strategies. Reactive approaches rely on real-time reallocation of staff, often guided by heuristics or predefined policies, while preventive approaches integrate flexibility into the initial schedule through buffers or backup assignments [45, 41].

More advanced strategies model the scheduling and rescheduling process as a two-stage stochastic optimization problem, where initial assignments are made under estimated conditions, followed by adjustments when new information arises. For example, Long et al. [47] introduced a multi-method rescheduling framework that allows reallocations both within and across departments or hospitals, showing improved adaptability to demand fluctuations. Such models highlight the importance of designing schedules that are both robust and responsive to dynamic healthcare environments.

1.2.6 Proactive-Reactive Strategies in Nurse Scheduling

Proactive scheduling strategies try to reduce problems before they happen by making the plan more flexible and strong from the start. These methods often use tools such as predictive models, robust optimization, and stochastic programming to deal with uncertainty, such as staff absences, changing patient demand, or delays in operations [40, 41]. In nurse scheduling, for example, proactive strategies may include assigning backup nurses, adding extra shifts as a buffer, or creating schedules where staff can easily replace one another [42]. These strategies help avoid last-minute changes, making the schedule more stable and improving staff satisfaction. Proactive planning is especially useful in healthcare, where unexpected absences can affect patient care. By including extra resources in the schedule from the beginning, hospitals can continue working smoothly without needing urgent changes [43].

Reactive strategies are used to make changes in real time when unexpected problems occur. These methods are important when proactive plans are not enough or when events are hard to predict [44]. Common reactive methods include updating the schedule using algorithms, calling in backup staff, or using simple rules to switch shifts [45]. For example, if a nurse is suddenly absent, a reactive strategy may involve calling other staff, changing who works where, or sharing tasks differently to cover the absence [41]. Although reactive strategies give quick answers, they can also lead to more work for managers, stress for staff, and less efficiency [46]. More and more, hospitals use hybrid systems that mix proactive and reactive methods like using data to start pre-planned backup actions to stay both prepared and flexible.

1.2.7 Review of Relevant Works

The problem of nurse scheduling and its dynamic counterpart, rescheduling under uncertainty, has been extensively studied in the literature. Numerous approaches have been proposed, ranging from exact optimization models and heuristics to hybrid methods integrating predictive analytics and machine learning. However, these works differ significantly in terms of their objectives, constraints, adaptability to uncertainty, and the healthcare systems they are designed for.

To contextualize and position our approach within the broader research landscape, we have selected a set of representative and influential studies. These studies have been chosen based on their methodological diversity, relevance to practical scheduling and rescheduling challenges, and their focus on uncertainty, flexibility, and optimization strategies. Our selection includes models applied to different healthcare systems, including but not limited to European contexts, and reflects a variety of solution techniques ranging from robust optimization to multi-objective approaches and heuristic frameworks.

In order to conduct a comprehensive yet focused review, we conducted an extensive literature search covering the last five years. While many studies have addressed nurse scheduling and rescheduling under uncertainty, only a subset of them directly aligns with our research context and goals. Therefore, we compiled a broader list of ten recent and relevant works, summarized in

Table 1.2, but chose to discuss in detail only three studies that are most pertinent to our framework. These selected studies stand out for their methodological rigor, their explicit treatment of uncertainty and sudden absences, and their alignment with practical healthcare scheduling needs. The complete list provides a broader view of current trends and contributions in the field, whereas the in-depth discussion highlights the most influential and comparable approaches to ours.

Table 1.3 provides an overview of the selected studies, including their main objectives, modeling techniques, and key contributions, and compares these works with our proposed framework, highlighting distinctions in handling uncertainty, the integration of predictive analytics, the use of overstaffing buffers, and the methodological innovation introduced through our two-stage approach.

This comparative analysis serves not only to underline the novelty and applicability of our framework, but also to identify current gaps in the literature. In particular, we emphasize the lack of real-time, data-driven, and low-disruption strategies that address nurse absences proactively while ensuring minimal perturbation of baseline schedules. By addressing these aspects, our work aims to contribute meaningfully to both theoretical and practical advances in the field.

Authors	Year	Title	Methodology
Otero-Caicedo et al.	2023	Preventive-Reactive Nurse Scheduling with Absenteeism	Multi-objective MILP solved by NSGA-II
Nagayoshi & Tamaki	2023	A Dynamic Nurse Scheduling Using Reinforcement Learning	Reinforcement Learning
Johansen et al.	2023	Nurse Scheduling & Rescheduling : Combining Optimization with Machine Learning-Driven Demand Predictions	MILP optimization + ML prediction (NN/DT) in rolling horizon
de Greef	2022	A Two-Stage Nurse Scheduling for Residential Care in the Netherlands	Two-stage scheduling with practical constraints
Long et al.	2022	Nurse Rescheduling with Multiple Methods under Uncertainty	Stochastic Programming & Distributionally Robust Optimization
Muniyan et al.	2022	Artificial Bee Colony Algorithm with Nelder-Mead for NSP	Metaheuristic optimization
Sari et al.	2021	An Innovation Scheduling Program During Covid-19	Heuristic scheduling
Valdano et al.	2021	Reorganization of Nurse Scheduling Reduces Infection Risk	Epidemic modeling & scheduling adjustment
Schoenfelder et al.	2020	Nurse Scheduling with Quick-Response Methods	Heuristics and simulation
Nagayoshi & Tamaki	2021	A Dynamic Nurse Scheduling for Various Sudden Absences	Reinforcement Learning
Karpuz & Batun	2019	Scheduling and Rescheduling in ICUs Under Uncertainty	Two-stage stochastic integer programming

Table 1.2: Comparison of Selected Nurse Scheduling and Rescheduling Studies

Among the vast literature on nurse scheduling and rescheduling, we selected four representative studies by Karpuz & Batun (2019), Long et al. (2022), Otero-Caicedo et al. (2023), and Johansen et al. (2023) to highlight diverse methodological perspectives and better position our proposed approach. These works cover a wide spectrum: from stochastic and distributionally robust formulations, to multi-objective optimization under absenteeism, and hybrid frameworks combining optimization with machine learning. Each provides valuable insights on handling uncertainty, designing reactive mechanisms, and validating results on real or simulated data. By comparing our contribution against these benchmarks, we aim to clearly illustrate the originality and relevance of our solution in addressing nurse scheduling disruptions in a dynamic and data-driven context.

Work 1: Johansen et al. (2023) – Nurse Scheduling and Rescheduling: Combining Optimization with Machine Learning-Driven Demand Predictions

Johansen et al. (2023) introduced a hybrid approach combining optimal nurse scheduling with dynamic rescheduling, embedded within a rolling horizon simulation. Their framework includes two MILP models: the first generates an initial schedule based on a weighted multi-objective function (covering hours, preferences, skills, etc.), while the second performs localized rescheduling to cover absences that remain unresolved by a heuristic phase. A distinctive feature of the study is the integration of machine learning models (neural networks and decision trees) to predict demand, which guides proactive scheduling decisions. Tested on real hospital data, the approach improved robustness to absences, enhanced coverage, and minimized disruptions to planned assignments.

Work 2: Otero-Caicedo et al. (2023) – Preventive-Reactive Nurse Scheduling with Absenteeism

Otero-Caicedo et al. (2023) proposed a preventive-reactive scheduling framework that anticipates absenteeism while integrating nurse preferences. A multi-objective MILP was solved using NSGA-II, enabling backup nurse assignments and schedule optimization. Reactive rescheduling was tested using three policies under simulated absenteeism and refusal rates. In a Colombian hospital case study, their approach improved nurse satisfaction by 11% and reduced workload disparities by 76%. Policy 1 (using backup nurses with minimal schedule change) achieved the best balance of performance and nurse satisfaction. Implementation via a desktop application demonstrated significant time savings and practical feasibility.

Work 3: Long et al. (2022) – Nurse Rescheduling with Multiple Methods under Uncertainty

Long et al. (2022) explored the nurse rescheduling problem under uncertain patient demand using real data from hospitals in China. The authors developed two optimization models: a stochastic programming model (SM) using historical scenarios and a distributionally robust model (DRM) leveraging ambiguity sets to handle limited data reliability. The models integrated multiple rescheduling methods (e.g., inter-departmental and inter-hospital transfers) and were reformulated into tractable integer programs. Results showed that SM achieved a 78.71% cost reduction, while DRM offered more robustness under high uncertainty with a 38.92% cost reduction. The DRM was more resilient to demand fluctuations, highlighting its suitability for unstable environments.

Work 4: Karpuz & Batun (2019) – Scheduling and Rescheduling in ICUs Under Uncertainty

Karpuz and Batun (2019) tackled the challenge of integrated scheduling and rescheduling in ICU settings, modeling the problem as a two-stage stochastic integer program. Monthly shift schedules formed the first stage, while daily rescheduling due to uncertain demand constituted the second. Using an AR(1) time series model to generate demand scenarios, they tested various cases with CPLEX. Their results showed that stochastic modeling provided substantial improvements: VSS ranged from 0.3% to 9.9%, while EVPI reached up to 89.8%, underscoring the value of uncertainty-aware planning. The study also explored cost sensitivity and proposed heuristics for low- and high-demand situations.

1.2.8 Our Contribution

This thesis is inspired by the work of Anne-Sofie Johansen, Bendik Nag, and Herborg Hermansen Tveit, from their 2023 Master's thesis at NTNU. They proposed a smart approach to nurse scheduling by combining optimization techniques with machine learning to predict demand. Their model was a strong starting point for handling uncertainty in nurse planning.

In our work, we start from their model and make several changes to better fit our goals. We modify the structure of the objective function by replacing the lexicographic method with a weighted-sum approach. This makes the model easier to use and much faster to solve, which is important for real-time or daily scheduling situations.

The main contribution of this thesis is a two-stage framework for dealing with unexpected absences in a smart and efficient way. In the first stage, we add a data-driven component that analyzes past absence data to calculate criticality scores for each shift. These scores show which periods are more at risk. Then, we adjust the model so that it gives more importance to having extra staff during these critical shifts. This helps make the schedule more stable from the beginning.

In the second stage, we design a simple heuristic that creates a buffer section in the schedule. This buffer is made up of additional nurses who are available to cover future absences. These nurses are chosen based on their skills, so they can replace others if necessary. When an absence occurs, the heuristic quickly checks if someone from the buffer can cover it. If so, the replacement is done immediately, without changing the rest of the schedule or running a new optimization.

Our results show that this heuristic can solve many absences on its own, with no cost and no disruption. By combining smart predictions with flexible planning, this two-step method brings a practical and effective solution for managing uncertainty in nurse scheduling.

Conclusion

In summary, this chapter has reviewed the foundational theories and recent advances in nurse scheduling, highlighting the shift toward data-driven and uncertainty-aware approaches. Despite progress, existing models often lack practical mechanisms to proactively manage unexpected absences with minimal disruption. Our proposed two-stage framework addresses this gap by integrating predictive analytics with operational flexibility, offering a promising solution for real-world healthcare planning.

Criteria	**Karpuz & Batun (2019)**	**Long et al. (2022)**	**Otero-Caicedo et al. (2023)**	**Johansen et al. (2023)**	**This Thesis (2025)**
Problem Focus	Scheduling & daily rescheduling (ICU)	Rescheduling under uncertain demand	Scheduling + Absenteeism Rescheduling	Combined scheduling-rescheduling with ML predictions	Scheduling + rescheduling under absences
Model Type	Two-stage stochastic program	Stochastic & Distributionally Robust Optimization	MILP + NSGA-II + Simulation	Multi-objective MILP + Neural Networks	Weighted MILP + criticality scores + heuristic
Uncertainty Handling	Time-series AR(1) model	Ambiguity sets & demand scenarios	Absenteeism simulation (Monte Carlo)	Markov chains for absences + ML demand forecasting	Historical data (critical shift scores)
Rescheduling Level	On-call, overtime/undertime	Inter-hospital & Inter-dept.	Backup policy + daily change	Cross-section buffers + flexible assignments	Proactive buffer + reactive heuristic
Real Data	×	✓	✓	✓(Cardiology clinic)	✓
Key Results / Contribution	VSS up to 9.9%, EVPI up to 89.8%	DRM robust to uncertainty; 78.71% cost reduction	Improved satisfaction (↑11%)	ML reduced costs by 15% vs baselines	2-stage framework: buffer section + fast rescheduling

Table 1.3: Comparative analysis of nurse scheduling and rescheduling approaches (chronological order)

Chapter 2

Methodology

Introduction

In this chapter, we present the baseline model for optimizing nurse scheduling, which serves as the basis of our approach. This model establishes the core structure, constraints, and initial objective formulation. We then introduce a series of enhancements that aim to improve the efficiency and robustness of the model in the face of uncertainty. These include a weighted-sum objective function that enables flexible prioritization among multiple goals, a criticality scoring function that identifies high-risk shifts based on historical absence trends, and a heuristic procedure for constructing a buffer section of surplus nurses, allowing for instant, hierarchy-aware substitutions before resorting to full rescheduling.

The analysis of absence data used in the criticality scoring function is detailed in the next chapter, Computational study. The figure below illustrates the overall methodology developed in this work, including the integration of proactive and reactive scheduling strategies within a unified two-stage framework.

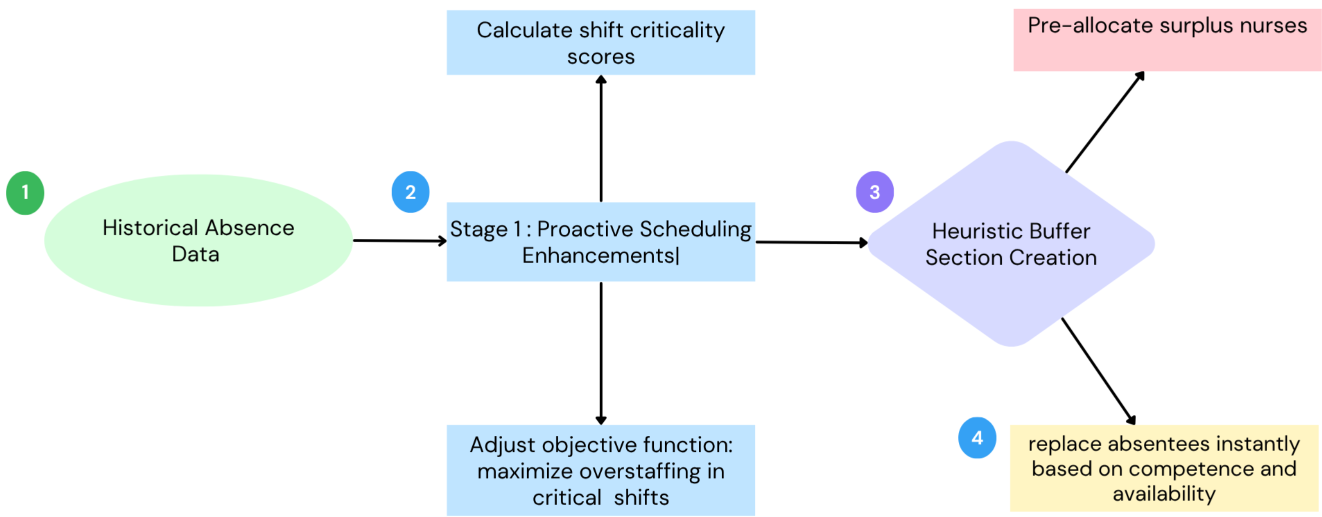

Figure 2.1: The Methodological framework

2.1 Nurse Scheduling Model

The optimization model presented in this section is based on the formulation developed by Johansen, Nag, and Tveit [24]. Their work, titled *Nurse Scheduling and Rescheduling: Combining Optimization with Machine Learning-Driven Demand Predictions*, provides a comprehensive mixed-integer programming approach for solving complex nurse scheduling problems. The model has been adapted to fit the context and data of the current study, while preserving its core structure and constraint logic.

The model aims to generate feasible nurse schedules that satisfy a wide range of operational constraints while optimizing multiple objectives such as demand coverage and contract compliance. The section begins by defining the relevant indices, sets, parameters, and decision variables. This is followed by the mathematical formulation of the objective functions and the constraints that govern the scheduling rules. Each component of the model is presented in detail to ensure clarity and reproducibility. Finally, the complete list of decision variables is declared, complete with the definition of the model.

Objective	**Description**
z_1	Minimize the number of understaffed shifts relative to the demand
z_2	Minimize total weekly deviation from contracted working hours for all nurses
z_3	Maximize the fair distribution of overstaffed shifts
z_4	Minimize the number of violations of individual nurse preferences
z_5	Minimize deviations from the required experience levels and the number of specialized nurses per shift

Table 2.1: Objective functions

Index	**Description**
$n \in \mathcal{N}$	Nurse
$q \in \mathcal{Q}$	Qualification level
$e \in \mathcal{E}$	Experience level
$u \in \mathcal{U}$	Unit or ward
$s \in \mathcal{S}$	Shift (e.g., morning, afternoon, night)
$d \in \mathcal{D}$	Day
$w \in \mathcal{W}$	Week

Table 2.2: Index

Sets	Description
$\mathcal{U}$	Set of hospital units or departments (e.g., $\mathcal{U} = \{1, 2, 3\}$)
$\mathcal{Q}$	Set of professional qualification levels (e.g., $\mathcal{Q} = \{$ AN: assistant nurse, N: nurse, SN: sp
$\mathcal{E}$	Set of experience categories assigned to nurses (e.g., $\mathcal{E} = \{\text{Junior}, \text{Mid-level}, \text{Senior}\}$)
$\mathcal{N}$	Complete set of nurses in the planning period
$\mathcal{N}^{(q)} \subseteq \mathcal{N}$	Subset of nurses with competence level $c \in \mathcal{C}$
$\mathcal{N}^{(e)} \subseteq \mathcal{N}$	Subset of nurses with experience level $e \in \mathcal{E}$
$\mathcal{N}^{(u)} \subseteq \mathcal{N}$	Subset of nurses assigned to unit $u \in \mathcal{U}$
$\mathcal{W}$	Set of weeks in the planning horizon
$\mathcal{D}$	Set of all days in the planning period
$\mathcal{D}_w \subseteq \mathcal{D}$	Days corresponding to week $w \in \mathcal{W}$
$\mathcal{D}^{\text{sun}} \subseteq \mathcal{D}$	Set of Sundays within the planning period
$\mathcal{S}$	Set of shift types (e.g., $\mathcal{S} = \{$D, E, N, F, F1$\}$)
$\mathcal{S}^{\text{work}} \subseteq \mathcal{S}$	Subset of working shifts (e.g., $\{$D, E, N$\}$)
$\mathcal{S}^{\text{off}} \subseteq \mathcal{S}$	Subset of non-working shifts (e.g., $\{$F, F1$\}$)

Table 2.3: List of sets

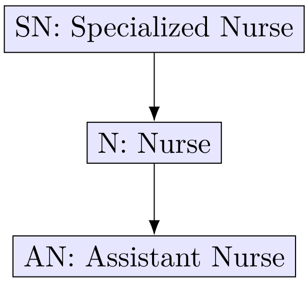

Figure 2.2: Hierarchy of qualification levels

In the model, the qualification levels are structured hierarchically. A **Specialized Nurse (SN)** possesses the highest qualification and can perform any task assigned to a **Nurse (N)** or an **Assistant Nurse (AN)**. Similarly, a **Nurse (N)** can cover tasks of an **Assistant Nurse (AN)**. This hierarchical substitution principle is essential for both scheduling and rescheduling phases, as it allows flexible reallocation of personnel based on competence levels when absences occur or shifts need to be reinforced.

Parameter	Description
D^{min}_{usd}	Minimum staffing requirement in unit u for shift s on day d
D^{avg}_{usd}	Average historical demand in unit u for shift s on day d
D^{exp}_{usd}	Preferred number of nurses with experience level e in unit b, shift s, day d
D^{AN}_{usd}	Maximum number of assistant nurses needed in unit b, shift s, day d
D^{SN}_{usd}	Desired number of senior nurses in unit b, shift s, day d
M^{day}	Maximum allowed number of consecutive workdays
M^{night}	Maximum allowed number of consecutive night shifts
H^{max}	Maximum total work hours allowed per week
H	Standard weekly hours for a full-time contract
H_s	Duration (in hours) of shift s
I^{rec}	Minimum interval between working weekends
C_n	Contracted working percentage for nurse n
$\overline{F}$	Upper limit on deviation from contracted weekly hours
$\underline{F}$	Lower limit on deviation from contracted weekly hours
K	Total number of weeks in the planning horizon
I^n_{usd}	1 if nurse n wants to avoid working shift s on day d in unit u, 0 otherwise

Table 2.4: List of parameters

Auxiliary Variables	Description
δ^{H-}_{nw}	Hours missing from the contractual workload for nurse n during week w
δ^{H+}_{nw}	Hours exceeding the contractual workload for nurse n during week w
δ^{SN-}_{usd}	Shortage of specialized nurses in unit u during shift s on day d
δ^{E}_{usd}	Shortage of nurses with experience level e in unit u, shift s, day d
δ^{D-}_{usd}	Staffing shortfall vs. average demand in unit u, shift s, day d
δ^{D+}_{usd}	Staffing surplus vs. average demand in unit u, shift s, day d

Table 2.5: Auxiliary Variables

Decision Variables

$$x_{nust} = \begin{cases} 1, & \text{if nurse } n \text{ in section } u \text{ is scheduled for shift } s \text{ on day } d \\ 0, & \text{otherwise} \end{cases}$$

$$\alpha_{ust} = \begin{cases} 1, & \text{if there is overstaffing in section } u \text{ on shift } s \text{ on day } d \\ 0, & \text{otherwise} \end{cases}$$

Objective functions

$$z_1 = \sum_{u \in U} \sum_{s \in S} \sum_{d \in D} \delta^{D-}_{usd} \tag{2.1}$$

$$z_2 = \sum_{n\in N}\sum_{w\in W}\left(\delta^{H-}_{nw} + \delta^{H+}_{nw}\right) \tag{2.2}$$

$$z_3 = \sum_{u\in U}\sum_{s\in S}\sum_{d\in D}\alpha_{usd} \tag{2.3}$$

$$z_4 = \sum_{n\in N}\sum_{u\in U}\sum_{s\in S}\sum_{d\in D} I^n_{usd}\cdot x_{nust} \tag{2.4}$$

$$z_5 = \sum_{u\in U}\sum_{s\in S}\sum_{d\in D}\left(\delta^{E}_{usd} + \delta^{SN-}_{usd}\right) \tag{2.5}$$

$$\min \sum \left(w_1 z_1 + w_2 z_2 - w_3 z_3 + w_4 z_4 + w_5 z_5\right) \tag{2.6}$$

Objective Related Constraints

$$\sum_{n\in\mathcal{N}} x_{nusd} = D^{\text{avg}}_{usd} - \delta^{D-}_{usd} + \delta^{D+}_{usd} \quad \forall u\in\mathcal{U},\ s\in\mathcal{S}^{\text{work}},\ d\in\mathcal{D} \tag{2.7}$$

(2.7) measures the deviation from the average demand by allowing both understaffing and overstaffing

$$\sum_{u\in\mathcal{U}}\sum_{s\in\mathcal{S}}\sum_{d\in\mathcal{D}_w} H_s\, x_{nusd} = C_n H + \delta^{H-}_{nw} - \delta^{H+}_{nw} \quad \forall n\in\mathcal{N},\ \forall w\in\mathcal{W} \tag{2.8}$$

(2.8) measures the deviations from contracted hours.

$$\sum_{n\in\mathcal{N}_e} x_{nusd} \geq D_{qusd} - \delta^{E}_{usd} \qquad e\in\mathcal{E}, u\in\mathcal{U}, s\in\mathcal{S}^{\text{work}}, d\in\mathcal{D} \tag{2.9}$$

(2.9) measures the deficit from desired demand for experience.

$$\sum_{n\in\mathcal{N}_{SN}} x_{nusd} \geq D^{SN}_{usd} - \delta^{SN-}_{usd} \qquad u\in\mathcal{U}, s\in\mathcal{S}^{\text{work}}, d\in\mathcal{D} \tag{2.10}$$

(2.10) measures the deficit from desired demand for competence.

$$\sum_{n\in\mathcal{N}(\text{AN})} x_{nusd} \leq \overline{D}^{\text{AN}}_{usd} \qquad \forall u\in\mathcal{U}, s\in\mathcal{S}^{\text{work}}, d\in\mathcal{D} \tag{2.11}$$

(2.11) limits the number of assistant nurses assigned to each shift.

$$\sum_{n\in\mathcal{N}} x_{nuNd} \leq D^{uNd}_{\min} + 3 \quad \forall u\in\mathcal{U},\ d\in\mathcal{D} \tag{2.12}$$

(2.12) permits night shift staffing to exceed the minimum requirement $D_{\min}$ by up to 3 nurses when necessary.

$$x_{n1sd} = 0 \qquad \forall n\in\mathcal{N}\setminus\mathcal{N}(u=1), s\in\mathcal{S}, d\in\mathcal{D} \tag{2.13}$$

$$x_{n2sd} = 0 \qquad \forall n\in\mathcal{N}\setminus\mathcal{N}(u=2), s\in\mathcal{S}, d\in\mathcal{D} \tag{2.14}$$

$$x_{n3sd} = 0 \qquad \forall n\in\mathcal{N}\setminus\mathcal{N}(u=3), s\in\mathcal{S}, d\in\mathcal{D} \tag{2.15}$$

Constraints (2.13)–(2.15) ensure nurses cannot be scheduled outside their assigned units ($\mathcal{N}(u)$ denotes nurses assigned to unit u).

Legislative Constraints

$$\sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}} x_{nusd} = 1 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.16}$$

(2.16) ensures each nurse is assigned exactly one shift per day.

$$\sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}^{\text{work}}} \sum_{d \in \mathcal{D}_w} H_s x_{nusd} \leq \overline{L} \quad \forall n \in \mathcal{N}, w \in \mathcal{W} \tag{2.17}$$

(2.17) limits weekly working hours for each nurse.

$$\underline{F} \sum_{k \in \mathcal{K}} C_n H \leq \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}^{\text{work}}} \sum_{d \in \mathcal{D}} H_s x_{nusd} \leq \overline{F} \sum_{k \in \mathcal{K}} C_n H \quad \forall n \in \mathcal{N} \tag{2.18}$$

(2.18) defines the contractual hours interval for each nurse.

$$\sum_{d'=d}^{d+\overline{M}^{\text{day}}} \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}^{\text{work}}} x_{nusd'} \leq \overline{M}^{\text{day}} \quad \forall n \in \mathcal{N}, d \in \{1, 2, ..., |\mathcal{D}| - \overline{M}^{\text{day}}\} \tag{2.19}$$

(2.19) limits consecutive working days.

$$\sum_{d'=d}^{d+\overline{M}^{\text{night}}} \sum_{u \in \mathcal{U}} x_{nu\text{N}d'} \leq \overline{M}^{\text{night}} \quad \forall n \in \mathcal{N}, d \in \{1, 2, ..., |\mathcal{D}| - \overline{M}^{\text{night}}\} \tag{2.20}$$

(2.20) limits consecutive night shifts.

Weekend Constraints

$$\sum_{w'=0}^{W-1} \sum_{u \in \mathcal{U}} \sum_{s \in \mathcal{S}^{\text{work}}} x_{nus(d+7w')} = 1 \quad \forall n \in \mathcal{N}, d \in \mathcal{D}^{\text{sun}} \tag{2.21}$$

(2.21) ensures each nurse works exactly one weekend every W weeks.

$$\sum_{u \in \mathcal{U}} \Big(x_{nuDd} - x_{nuE(d-1)} \Big) = 0 \quad \forall n \in \mathcal{N}, d \in \mathcal{D}^{\text{sun}} \tag{2.22}$$

$$\sum_{u \in \mathcal{U}} \Big(x_{nuEd} - x_{nuD(d-1)} \Big) = 0 \quad \forall n \in \mathcal{N}, d \in \mathcal{D}^{\text{sun}} \tag{2.23}$$

$$\sum_{u \in \mathcal{U}} \Big(x_{nuNd} - x_{nuN(d-1)} \Big) = 0 \quad \forall n \in \mathcal{N}, d \in \mathcal{D}^{\text{sun}} \tag{2.24}$$

Constraints (2.22)-(2.24) enforce legal weekend shift patterns.

Rest Regulations

$$\sum_{u \in \mathcal{U}} \Big(x_{nuNd} + x_{nuD(d+1)} \Big) \leq 1 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.25}$$

(2.25) prohibits night-to-morning shift transitions.

$$\sum_{u \in \mathcal{U}} \Big(x_{nuNd} + x_{nuE(d+1)} \Big) \leq 1 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.26}$$

(2.26) prohibits night-to-evening shift transitions.

$$\sum_{u \in \mathcal{U}} \left(x_{nuN(d-1)} + x_{nuF_1d} + x_{nuD(d+1)} \right) \leq 2 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.27}$$

$$\sum_{u \in \mathcal{U}} \left(x_{nuN(d-1)} + x_{nuF_1d} + x_{nuE(d+1)} \right) \leq 2 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.28}$$

$$\sum_{u \in \mathcal{U}} \left(x_{nuE(d-1)} + x_{nuF_1d} + x_{nuD(d+1)} \right) \leq 2 \quad \forall n \in \mathcal{N}, d \in \mathcal{D} \tag{2.29}$$

(2.27),(2.29) regulate weekly rest day (F_1) patterns.

$$\sum_{n \in \mathcal{N}} \sum_{u \in \mathcal{U}} \sum_{d \in \mathcal{D}_w} x_{nuF_1d} = 1 \quad \forall w \in \mathcal{W} \tag{2.30}$$

(2.30) ensures one rest day (F_1) per week per nurse.

$$\sum_{u \in \mathcal{U}} (x_{nuF1d-6} + x_{nuDd-5}) \leq 1 \quad \forall n \in \mathcal{N},\ d \in \mathcal{D}^{\text{sun}} \tag{2.31}$$

(2.31) prevents nurses from working both:

- An F1 (off-shift) on the last day before the weekend ($x_{nuF1d-6}$)
- A Day shift on the first day after the weekend (x_{nuDd-5})

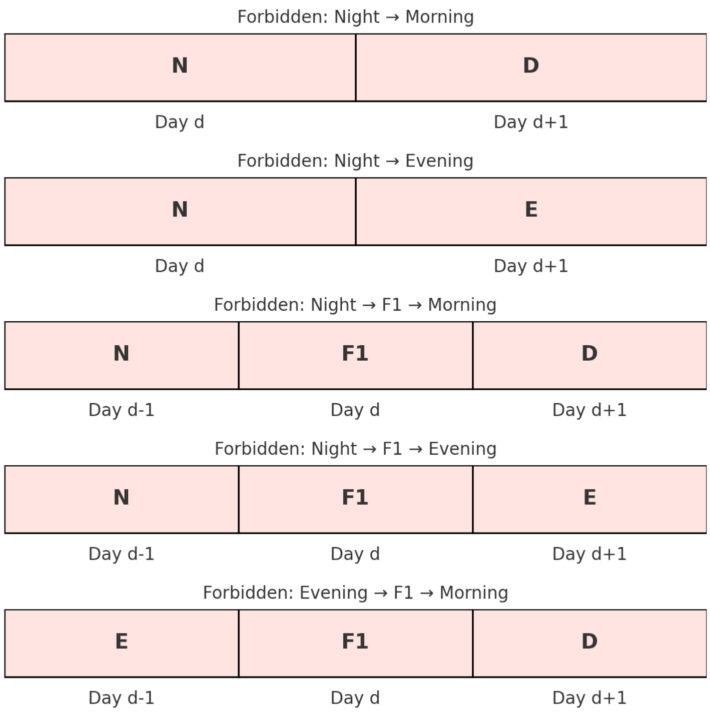

Figure 2.3: Examples of forbidden rest patterns involving night, F1, and day/evening shifts

Model Assumptions This model makes the following assumptions:

- Nurse availability and absence data are known or predicted prior to schedule generation.
- The planning horizon is segmented into days and fixed shift types (D, E, N).
- Each nurse is assigned exactly one shift per day and cannot be split between units.
- Absences are not anticipated within the optimization but handled in the second stage.
- The qualification hierarchy (SN > N > AN) is strictly respected in substitution.

2.2 Data-Driven Model

The base scheduling model presented above incorporates both understaffing and overstaffing through soft constraints. It also integrates a variety of operational, contractual, and regulatory rules to ensure the feasibility of the generated schedules. However, an important limitation is that it treats all shifts equally when allocating surplus staff. In other words, any allowed overstaffing is distributed uniformly throughout the planning horizon, without considering the varying levels of risk associated with different shifts.

In practice, not all shifts are equally exposed to disruptions. Certain combinations of weekday and shift type, such as Monday mornings or weekend nights, are statistically more prone to nurse absences. Ignoring this heterogeneity may result in inefficient use of surplus resources, as the model might assign additional staff where the risk of absence is low, while failing to reinforce the shifts that are statistically more likely to face staffing shortages.

Overstaffing, when used strategically, can serve as a powerful tool to absorb unplanned absences and reduce the need for disruptive rescheduling actions. However, its effectiveness depends on the ability to anticipate which shifts are at higher risk. To address this issue, we introduce a data-driven approach to identify what we define as **critical shifts**: those shifts that, based on historical data, have shown consistently higher levels of absenteeism.

Once identified, these critical shifts will be used to guide the optimization process. In the enhanced version of the model, overstaffing is no longer uniformly rewarded but instead becomes shift-specific encouraging surplus capacity where it is most likely to have a positive impact. This refinement aims to make the schedule not only feasible and compliant but also more robust in the face of real-world uncertainty.

2.2.1 Data-Driven Identification of Critical Shifts

This identification is based on the historical absence records for each hospital section. For each record, we compute the actual absence levels per shift (*Morning*, *Evening*, *Night*) by comparing the planned and actual number of nurses. Negative absence values are discarded by applying a non-negativity constraint.

For each shift and weekday, we aggregate the data to compute:

- **Criticality score:** a weighted average of observed absence levels. This score reflects how critical a shift is in terms of expected staffing shortfalls.

A shift is classified as **critical** based on its criticality score. The output is a ranked list of shifts that decrease in criticality.

The pseudocode below summarizes the algorithm used:

Algorithm 1 Identify Critical Shifts from Historical Absence Data

Input: Historical shift-level absence data for each hospital section

Output: Sorted list of shifts ranked by criticality score

foreach *hospital section in the dataset* **do**

Import historical records containing planned and actual staffing levels per shift; For each date entry, determine the corresponding weekday (Monday, Tuesday, ...);

ruby Copier Modifier Compute the number of absent nurses per shift using non-negativity:

- Morning_absences $\leftarrow \max(0, \texttt{Planned\_Morning} - \texttt{Actual\_Morning})$
- Evening_absences $\leftarrow \max(0, \texttt{Planned\_Evening} - \texttt{Actual\_Evening})$
- Night_absences $\leftarrow \max(0, \texttt{Planned\_Night} - \texttt{Actual\_Night})$

foreach *shift* $\in$ *{Morning, Evening, Night}* **do**

Aggregate absence data by weekday For each weekday:

- Count the number of days with different absence levels
- Let n_d be the total number of days observed for that shift and weekday
- Let a_i be the number of days with i absences ($i \in \mathbb{N}$)

Compute the **criticality score** as:

$$\text{Criticality Score} = \frac{\sum_{i=1}^{k} i \cdot a_i}{n_d}$$

where k is the maximum number of observed absences

Add shift, weekday, section, and criticality score to results list

Sort the list of shifts by descending criticality score; **return** *Ranked list of critical shifts*

Identified critical shifts serve as the foundation for improving the model of optimization of nurse scheduling. In the next section, we explain how these scores inform a criticality-aware allocation strategy and guide the optimization process toward resilient staffing decisions.

2.2.2 Integrating Criticality into the Optimization Model

After identifying the shifts that are most often affected by absences, the next step is to include this information in the scheduling model. The aim is to move from a uniform way of allocating surplus nurses to a more focused approach that gives priority to overstaffing where it can reduce the risk of future absences.

For this purpose, we use the criticality scores that were calculated earlier. These scores represent the level of risk for each shift and help the model assign additional staff to the most vulnerable periods. The effect of this approach on staffing performance will be analyzed in the *Computational Study* chapter.

The enhancement consists in replacing the original static surplus reward term w_3 with a more adaptive mechanism based on two components:

- A **binary reward** $\lambda_1 \cdot \alpha_{usd}$, activated whenever at least one surplus nurse ($\alpha_{usd} = 1$) is assigned to a shift.

- A **continuous reward** $\lambda_2 \cdot \delta_{usd}^{D+}$, proportional to the amount of overstaffing, weighted by the shift's criticality score.

These two components are combined into the enhanced surplus objective function:

Note on notation. The term z_3 was originally defined as a basic surplus tracking component. In this section, it is replaced by z_3^{enhanced} to reflect critical-aware overstaffing. Unless otherwise stated, all later formulations refer to this improved version.

$$z_3^{\text{enhanced}} = \sum_{(u,s,d)\in\mathcal{C}} \left(\lambda_1 \alpha_{usd} + \lambda_2 \delta_{usd}^{D+}\right) \cdot \text{criticality}_{usd} \tag{2.32}$$

In this formulation, $\mathcal{C}$ denotes the set of shifts classified as critical, and $\text{criticality}_{usd})isthenormalizedabsencerisk$ shift s, and day d. Because the criticality score reflects both the severity and frequency of past absences, it implicitly prioritizes the shifts with the highest vulnerability.

To prevent excessive overstaffing, we impose a constraint that caps the maximum surplus assigned to any shift.

$$\delta_{usd}^{D+} \leq \overline{M}^{\text{surp}} \quad \forall u \in \mathcal{U},\ s \in \mathcal{S}^{\text{work}},\ d \in \mathcal{D} \tag{2.33}$$

This constraint ensures a balance between flexibility and cost-efficiency while keeping surplus staffing focused on high-risk periods.

The enhanced surplus reward is then incorporated into the global objective function, which maintains all other components and constraints from the base model:

$$\min \sum \left(w_1 z_1 + w_2 z_2 - z_3^{\text{enhanced}} + w_4 z_4 + w_5 z_5\right) \tag{2.34}$$

In summary, this enhancement transforms the base model into a more robust and data-aware system. By leveraging absence history through criticality scores, it directs surplus capacity to the most sensitive shifts, reinforcing resilience without compromising feasibility or operational constraints.

2.3 Heuristic: Nurse Reassignment with Overflow Unit

Since maximizing overstaffing in critical shifts is a key objective in our scheduling model, we also reflect this principle in the absence management strategy. To address unplanned absences in a simple yet effective way, we develop a post-optimization heuristic called *Nurse Reassignment with Overflow Unit*. This approach allows local adjustments to be made to the initial schedule without re-running the full optimization process, which can be time-consuming and computationally expensive.

The heuristic operates in two phases. In the first phase, it scans the schedule to detect surplus staff. For each unit, shift, and day, it compares the current number of nurses assigned to the target demand. If a surplus is detected, the algorithm removes the least qualified nurses from their original assignments and reassigns them to a dedicated overflow unit. This overflow unit acts as a flexible pool of standby nurses, available for reassignment when needed.

In the second phase, the algorithm processes the list of unplanned absences. For each absent nurse, it checks whether the remaining staff still meet the required demand. If not, it looks into the overflow unit for candidates who are available in the same shift and day. The replacement nurse must have a qualification level equal to or higher than that of the absent nurse. Among eligible candidates, the algorithm selects the one with the lowest acceptable qualification to preserve higher-skilled staff for future reassignments if needed.

This method enables quick and targeted corrections to the schedule. By relying only on pre-allocated surplus capacity and basic qualification checks, the heuristic avoids major disruptions and maintains operational feasibility. Moreover, since the reassignments are performed locally, the computation time remains negligible.

The algorithm also tracks statistics such as the number of absences covered, the number of surplus nurses reassigned, and the average qualification gap between absent and replacement nurses. These indicators provide useful feedback on the effectiveness of the heuristic and can guide future adjustments to buffer size or qualification allocation policies.

The full pseudo-code and the logic of this heuristic is summarized below:

Algorithm 2 Simple Reassignment Heuristic with Overflow and Absence Management (without average gap)

Input: Initial nurse assignment $X(n, u, s, d)$,
Desired demand $D_{\text{des}}[u, s, d]$,
Nurses grouped by qualification levels N_q,
Absence indicator $A(n, u, s, d)$ (1 if absent, 0 otherwise)
Output: Updated nurse assignment X,
Statistics on reassignments and covered absences
Initialize qualification levels:
 level$(n) \leftarrow i$ if $n \in N_q[i]$
 Initialize counters: surplus_reassigned $\leftarrow 0$, absences_covered $\leftarrow 0$
foreach $u \in U \setminus \{overflow\}$, $s \in S^{work}$, $d \in D$ **do**
  Compute current coverage:
   $C \leftarrow \sum_n X(n, u, s, d)$
   Surplus: surplus $\leftarrow C - D_{\text{des}}[u, s, d]$
  if *surplus* > 0 **then**
   Sort nurses assigned to (u, s, d) by ascending qualification level
    foreach *lowest qualification nurse n_k among surplus* **do**
     Remove nurse from unit b: $X(n_k, u, s, d) \leftarrow 0$
      Assign nurse to overflow unit: $X(n_k, \text{overflow}, s, d) \leftarrow 1$
      surplus_reassigned $\leftarrow$ surplus_reassigned $+ 1$
foreach *absent nurse (n, u, s, d) with $A(n, u, s, d) = 1$* **do**
  Remove absent nurse: $X(n, u, s, d) \leftarrow 0$
   if $\sum_m X(m, u, s, d) \geq D_{des}[u, s, d]$ **then**
    continue ▷ Demand already met
  Find candidates in overflow unit at (s, d) with qualification $\geq$ level(n)
   if *candidates exist* **then**
    Select candidate $m^\star$ with minimal qualification level
     Reassign candidate: $X(m^\star, \text{overflow}, s, t) \leftarrow 0$, $X(m^\star, u, s, d) \leftarrow 1$
     absences_covered $\leftarrow$ absences_covered $+ 1$
return *Updated assignment X and statistics {surplus_reassigned, absences_covered}*

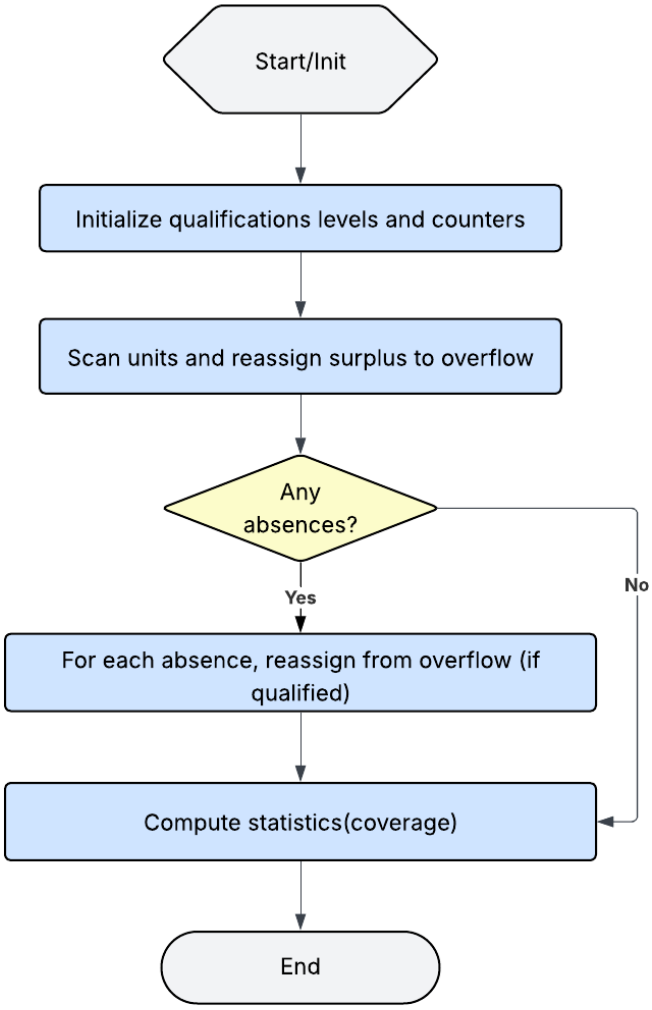

Figure 2.4: Framework connection

Integration of Scheduling and Rescheduling within a Simulation-Based Framework

The proposed methodology adopts a simulation-based modular approach structured around a two-phase decision process. This framework is designed to balance proactive planning with reactive adaptability, enabling robust management of the nurse workforce under uncertainty.

Phase 1: Optimized Initial Scheduling

At the beginning of each simulation cycle, a mixed-integer linear programming model is executed to generate a high-quality schedule over a fixed planning horizon. This optimization considers a comprehensive set of operational constraints, including staffing requirements, contractual obligations, rest regulations, and hierarchical qualification rules. Importantly, the model integrates a criticality-aware overstaffing strategy, leveraging historical absence patterns to prioritize staffing in high-risk shifts. The output of this phase is a feasible, optimized plan that incorporates the targeted surplus capacity in the form of a virtual *overflow unit*.

Phase 2: Daily Rescheduling through Heuristic Adjustment

Once the initial schedule is generated, the system enters a rolling-horizon simulation loop. Each day, new absences are simulated. In response, a rescheduling heuristic is triggered to locally reassign nurses. This heuristic searches the overflow unit for qualified surplus nurses available on the current day and reassigns them to uncovered shifts while strictly respecting hierarchical substitution rules (i.e., a specialized nurse can replace a regular nurse, but not vice versa). The process is efficient, non-disruptive, and avoids recomputing the full schedule.

This architecture ensures a continuous interaction between long-term planning and short-term responsiveness. Rather than relying on external predictors, the method leverages historical absence patterns (through criticality scores) and internal flexibility (through the overflow buffer) to dynamically absorb daily disruptions. The simulation framework supports two approaches to generate absence scenarios: initially, it uses real historical absence data to evaluate the model in realistic conditions; subsequently, it can integrate a predictive machine learning model trained on historical patterns to simulate plausible future absences. As a result, the framework remains simple, adaptable, and computationally tractable qualities that are essential for real-world hospital deployment.

Figure 2.5 illustrates the conceptual integration of these two phases within the rolling simulation loop.

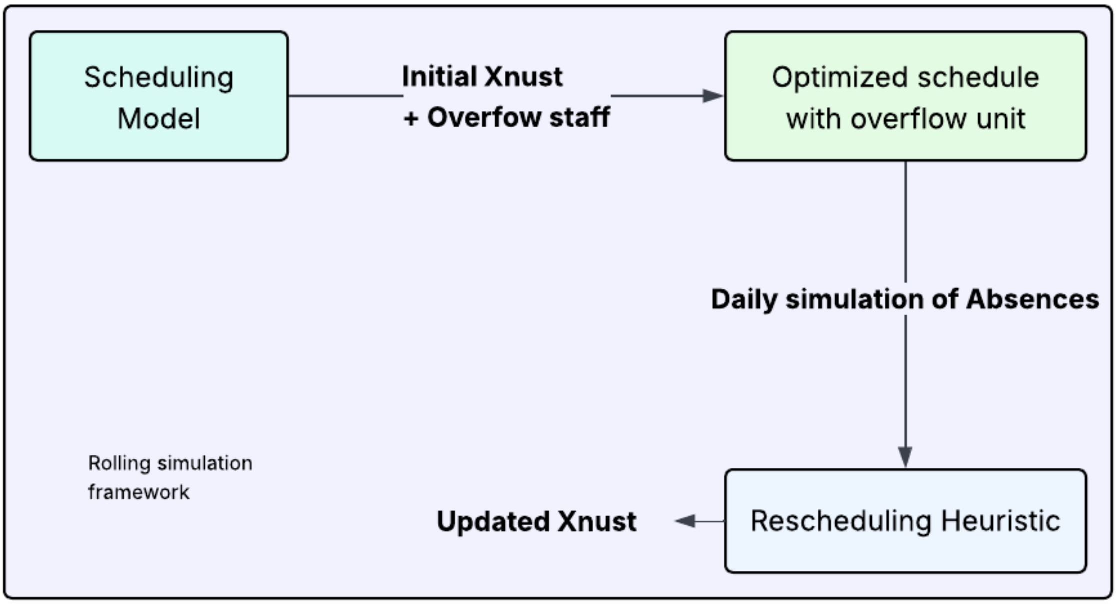

Figure 2.5: Integration of Scheduling and Rescheduling within a Simulation- Based Framework

Conclusion

This chapter introduced a two-phase nurse scheduling framework that combines robust optimization with a lightweight rescheduling heuristic. The core model incorporates operational and regulatory constraints, while a criticality-aware objective function directs surplus capacity toward high-risk shifts.

To handle daily disruptions, a heuristic uses an overflow unit of pre-allocated nurses for rapid, hierarchy-compliant reassignments. In the next chapter, we present the results of the different tests conducted to evaluate the effectiveness and robustness of this framework under various absence scenarios.

Chapter 3

Computational Study

Introduction

This chapter builds on the models developed in the previous section by moving from design to experimentation. We aim to assess the performance of the proposed nurse scheduling and rescheduling solutions through a series of computational tests. These experiments are based on synthetic data that reflect real hospital conditions, including staffing constraints, preferences, and absenteeism. We describe the different steps of our approach, from data initialization to schedule evaluation, and analyze the results to validate the quality, robustness, and practical relevance of the proposed methods.

3.1 Test Environment

All models and analytical procedures were implemented in Python. The optimization models were solved using the commercial IBM ILOG CPLEX Optimizer. All experiments were performed on a personal computer with moderate computational capacity. The hardware and software specifications of the test environment are summarized in Table 3.1.

Specification	**Details**
Processor	Intel(R) Core(TM) i7-7500U
Cores / Frequency	2 / 2.70–2.90GHz
Operating System	Windows 10 Home, 64-bit
RAM	8 GB
Python version	3.10.9
CPLEX version	22.1.1

Table 3.1: Details of software and hardware specifications

3.2 Data Generation

To evaluate the performance of our nurse scheduling and rescheduling models under realistic conditions, we generated synthetic data that simulate real-world hospital environments. This data includes nurse shift preferences and unplanned absences, both of which introduce significant complexity into the scheduling process. The goal is to simulate scenarios that reflect the actual variability and uncertainty faced by healthcare planners. In the following subsections, we describe the methods used to generate preference and absence data.

3.2.1 Nurse Preference Generation Method

To better reflect real-world behavior in our scheduling model, we created a method to simulate nurse preferences for different shifts. This method is based on historical data showing when nurses reported not wanting to work called "violations." These violations are classified by shift type: day ($ViolD$), evening ($ViolE$), and night ($ViolN$). A detailed analysis of the original preference data can be found in Section 3.7.

We first analyze how often these violations occur for each day of the week and for each type of shift. From this we calculate probabilities that indicate the likelihood that a nurse would not like a particular shift on a given day. For example, if many nurses tend to avoid night shifts on Mondays, our model assigns a higher probability to that combination of shift-day.

We also pay special attention to the weekends. Since nurses often prefer to avoid working both Saturday and Sunday, we calculate separate weekend-specific probabilities. These weekend preferences are smoothed: if a nurse dislikes working one weekend day, the model assumes that they are also likely to dislike the other. This generates more realistic patterns that reflect actual human behavior.

Using these probabilities, we construct a preference matrix for each nurse. This matrix, denoted as `Inst`, has one row for each day and one column for each shift type (day, evening, night). Each value in the matrix is binary: 1 means that the nurse does not want to work that shift on that day, and 0 otherwise. Uniform random numbers are generated and compared against the corresponding probability; if the random number is less than the probability, a 1 is assigned; otherwise, a 0. This procedure ensures that preferences are both data-driven and individualized (Hillier Lieberman, 2015).

In summary, our method uses past behavior to simulate realistic, personalized nurse preferences. These preferences are critical for testing and improving our scheduling algorithms, as they capture the complexity of human availability and satisfaction.

3.2.2 Absence Data

To introduce realistic disturbances into the planning process, we construct a dictionary of expected absences for each day, section, and shift. This dictionary serves as an input to the simulation framework and defines the number of staff expected to be absent at each point in time.

The structure is as follows:

$$\texttt{absences\_dict}[(d, u, s)] = n_{\text{abs}}$$

Where:

- d is the day index,
- u is the unit identifier,
- s is the shift (morning, evening, night),
- n_{abs} is the number of absentees expected for that combination.

This absence dictionary was generated using two approaches:

- **Empirical data:** Directly extracted from historical records, measuring the gap between planned and actual staffing.
- **Hurdle model prediction:** A two-part statistical model trained on features such as weekday, month, workload indicators, and public holidays. Estimate both the probability and the magnitude of absenteeism.

The resulting dictionary does not assign absences to specific individuals, but only specifies the expected number of absentees, which are drawn dynamically during the simulation.

3.3 Rolling Horizon Simulation Framework

To evaluate the behavior of the scheduling system under dynamic and uncertain conditions, we implemented a rolling horizon simulation framework. This framework simulates the evolution of the planning process day by day, incorporating unexpected absences and triggering reactive reassignment strategies.

Each day in the planning horizon is treated as a simulation step during which the following operations are executed sequentially:

1. **Planning extraction:** The daily schedule is extracted from the initial planning. This includes all nurse assignments for each section and shift.
2. **Absence sampling:** For each day d, section u, and shift s, a predefined number of absentees n_{abs} is drawn based on the dictionary `absences_dict`$[(d, u, s)]$ (Section 3.2.2). Among the nurses scheduled to work at (u, s, d), a random sample of size n_{abs} is selected without replacement. If fewer than n_{abs} nurses are available, all are marked absent. This ensures that absences are only assigned to staff who were effectively planned to work.
3. **Reassignment heuristic:** Once absences are identified, a reassignment heuristic is invoked to mitigate their impact. We use the *Nurse Reassignment with Overflow Unit* described in Section 2.3, which reallocates surplus staff from overstaffed units to an overflow unit and reassigns them to uncovered shifts based on qualification constraints. This approach allows for rapid and localized adjustment of the schedule without re-optimizing entire planning.
4. **Logging and evaluation:** Key indicators are recorded for each day, including the number of simulated absences, the number of successfully reassigned staff and the residual uncovered demand. These metrics are then aggregated to assess the robustness of planning and rescheduling strategies.

This rolling procedure allows us to simulate a realistic operational environment, where planning is constantly challenged by uncertainty. It also enables the evaluation of reactive strategies over an extended horizon, providing insights into the strengths and limitations of the scheduling system.

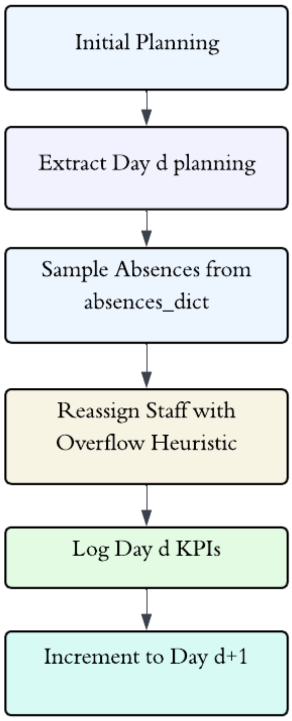

Figure 3.1: Rolling Horizon Simulation Process

3.4 Initialization of scheduling Parameters

To simulate a realistic nurse scheduling environment, the dataset was initialized with a diversified pool of nurses categorized by experience levels. Specifically, nurses were grouped into three categories according to their professional experience: junior, intermediate, and senior.

The planning horizon was set at four weeks, broken down into daily time steps and weekly partitions to align with hospital operational cycles. The shifts were categorized into five types, distinguishing between weekday shifts (S_W) and weekend shifts (S_O). The corresponding demand data, both minimum and desired staffing levels, was initialized for each combination of department, day, and shift, using multi-dimensional arrays D^{min}_{usd} and D^{avg}_{usd}. These were derived from domain-specific estimates and reflect typical hospital demand fluctuations during the week and weekends.

Each nurse was assigned an individual contractual workload, expressed as a percentage, which made up the array C_n. These values determine the expected working hours over the entire planning horizon, computed using the formula:

$$C_n[i] = K \times H \times \text{percentage}_i$$

where H represents the standard weekly workload in hours and K is the number of weeks in the planning horizon.

Parameter	Description	Value
M^{day}	Maximum allowed number of consecutive workdays	5
M^{night}	Maximum allowed number of consecutive night shifts	3
L	Maximum total work hours over the planning horizon (4 weeks)	48
H	Standard weekly hours for a full-time contract	35.5
H^{day}	Duration of a day shift in hours	7.5
H^{eve}	Duration of an evening shift in hours	7
H^{night}	Duration of a night shift in hours	10
I^{rec}	Minimum interval between working weekends (in weeks)	2
$\overline{F}$	Upper limit on deviation from contracted weekly hours	105%
$\underline{F}$	Lower limit on deviation from contracted weekly hours	95%
K	Total number of weeks in the planning horizon	4

Table 3.2: Model Parameters, Descriptions, and Values

3.5 Baseline Model Evaluation

To evaluate the effectiveness and scalability of the proposed multi-objective nurse scheduling model, we conducted a series of experiments on instances of increasing complexity. The evaluation began with a small instance of 10 nurses over a one-week planning horizon and was extended to larger, more realistic scenarios involving 104 nurses over one to four weeks. This progression allows for clearer and more precise visualization of the schedules and highlights how the model performs under different constraints.

In our analysis, we focus on comparing two main multi-objective optimization strategies: the weighted-sum (scalarization) approach and the lexicographic approach, both with and without slack variables. This comparison aims to reveal how each method balances trade-offs between conflicting objectives, and how the presence or absence of slack influences performance and feasibility.

Based on the results obtained from these experiments, we will identify the most suitable approach and use it to generate the final nurse planning solutions for larger instances.

The metrics considered are defined as follows:

- **Understaffing (z1)**: Number of understaffed shifts.
- **Hours deviation (z2)**: The total weekly hours deviated from the contracted hours.
- **Max deviation**: Maximum number of hours deviated for a single nurse.
- **Overstaffing (z3)**: Number of overstaffed shifts.
- **Preference (z4)**: Percentage of satisfied nurse preferences.
- **Experience violation (z5)**: Number of shifts that violate the required level of experience.
- **Competence violation**: Number of shifts that violate the required competence level.
- **Run Time**

3.5.1 Small Instance: 10 Nurses, 1 Week

This initial test case provides a detailed look at the behavior of the model in a simplified scenario. Table 3.3 presents the results obtained using the three optimization strategies: scalarization, lexicographic with slack, and lexicographic without slack.

Scalar Approach: $12Z_1 + 8Z_2 - 6Z_3 + 4Z_4 + 2Z_5$

Metric	**Lexicographical**		**Scalar-**
	With slack	**Without slack**	**approach**
Understaffing (x_1)	0.00	0.00	0.00
Hours deviation (x_2)	2.67	**1.67**	2.67
Max deviation	1.00	**0.5**	1.00
Overstaffing (x_3)	6.00	6.00	6.00
Preference (x_4)	86.47	86.47	86.47
Experience violation (x_5)	**1.00**	3.00	**1.00**
Competence violation	**2.00**	4.00	**2.00**
Run Time (s)	1.98	17.00	2.22

Table 3.3: Comparison of multi-objective approaches on an instance with 10 nurses and 1 unit

Nurses	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Work Contrat(%)
1	F (Sec 1)	F (Sec 1)	D (Sec 1)	F (Sec 1)	N (Sec 1)	F1 (Sec 1)	F (Sec 1)	50
11	F (Sec 1)	E (Sec 1)	F1 (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	E (Sec 1)	100
17	E (Sec 1)	D (Sec 1)	E (Sec 1)	D (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	100
24	F (Sec 1)	F1 (Sec 1)	E (Sec 1)	F (Sec 1)	F (Sec 1)	E (Sec 1)	D (Sec 1)	60
39	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	F (Sec 1)	50
53	N (Sec 1)	F (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	D (Sec 1)	E (Sec 1)	90
71	D (Sec 1)	F (Sec 1)	F (Sec 1)	D (Sec 1)	F1 (Sec 1)	N (Sec 1)	N (Sec 1)	100
94	E (Sec 1)	E (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	F1 (Sec 1)	F (Sec 1)	100
97	N (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	E (Sec 1)	F1 (Sec 1)	F (Sec 1)	75
103	F (Sec 1)	D (Sec 1)	N (Sec 1)	F (Sec 1)	F1 (Sec 1)	E (Sec 1)	D (Sec 1)	90

Shifts :
D: Day
E :Evening
N: Night

F1 : Weekly minimum consecutive rest
F: Off shift

Lexicographic Approach (with slack)

Figure 3.2: Weekly Planning for 10 Nurses(Lexicographic with slack)

Nurses	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Work Contrat(%)
1	D (Sec 1)	F (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	F1 (Sec 1)	F (Sec 1)	50
11	F (Sec 1)	E (Sec 1)	F1 (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	E (Sec 1)	100
17	E (Sec 1)	E (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	100
24	F (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	F (Sec 1)	60
39	F (Sec 1)	F (Sec 1)	D (Sec 1)	N (Sec 1)	F (Sec 1)	F1 (Sec 1)	F (Sec 1)	50
53	D (Sec 1)	D (Sec 1)	N (Sec 1)	F (Sec 1)	E (Sec 1)	F (Sec 1)	F1 (Sec 1)	90
71	F (Sec 1)	F (Sec 1)	D (Sec 1)	D (Sec 1)	F1 (Sec 1)	N (Sec 1)	N (Sec 1)	100
94	E (Sec 1)	D (Sec 1)	E (Sec 1)	E (Sec 1)	E (Sec 1)	F (Sec 1)	F1 (Sec 1)	100
97	E (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	N (Sec 1)	F1 (Sec 1)	F (Sec 1)	75
103	N (Sec 1)	F (Sec 1)	F (Sec 1)	F1 (Sec 1)	D (Sec 1)	E (Sec 1)	D (Sec 1)	90

Shifts :
D: Day
E :Evening
N: Night

F1 : Weekly minimum consecutive rest
F: Off shift

Scalar Approach (without slack)

Figure 3.3: Weekly Planning for 10 Nurses(Lexicographic without slack)

Nurses	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday	Work Contrat(%)
1	D (Sec 1)	F (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	F1 (Sec 1)	F (Sec 1)	50
11	F1 (Sec 1)	E (Sec 1)	F (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	E (Sec 1)	100
17	E (Sec 1)	D (Sec 1)	E (Sec 1)	D (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	100
24	D (Sec 1)	E (Sec 1)	E (Sec 1)	F (Sec 1)	F (Sec 1)	F (Sec 1)	F1 (Sec 1)	60
39	D (Sec 1)	F (Sec 1)	F (Sec 1)	F (Sec 1)	N (Sec 1)	F1 (Sec 1)	F (Sec 1)	50
53	F1 (Sec 1)	F (Sec 1)	D (Sec 1)	N (Sec 1)	F (Sec 1)	E (Sec 1)	D (Sec 1)	90
71	F (Sec 1)	D (Sec 1)	F (Sec 1)	F1 (Sec 1)	D (Sec 1)	N (Sec 1)	N (Sec 1)	100
94	E (Sec 1)	F (Sec 1)	E (Sec 1)	F1 (Sec 1)	E (Sec 1)	E (Sec 1)	D (Sec 1)	100
97	N (Sec 1)	N (Sec 1)	F (Sec 1)	F (Sec 1)	E (Sec 1)	F1 (Sec 1)	F (Sec 1)	75
103	E (Sec 1)	F (Sec 1)	D (Sec 1)	D (Sec 1)	N (Sec 1)	F1 (Sec 1)	F (Sec 1)	90

Shifts :
D: Day
E :Evening
N: Night

F1 : Weekly minimum consecutive rest
F: Off shift

Scalar Approach

Figure 3.4: Weekly Planning for 10 Nurses (scalar approach)

The comparison of multi-objective approaches reveals significant differences in performance metrics. The lexicographical method without slack achieves better results in hours deviation and max deviation compared to both the lexicographical method with slack and the scalar approach, but at the cost of substantially longer runtime. Both the lexicographical with slack and scalar approaches demonstrate superior performance in minimizing experience violations (1.00 vs. 3.00) and competence violations. All methods maintain identical results for understaffing, overstaffing and preference satisfaction. This suggests that while the lexicographical method without slack optimizes certain scheduling metrics better, it comes with significantly increased computational time, making the with-slack and scalar approaches more efficient for practical implementations where experience and competence constraints are prioritized.

To confirm these observations, further tests on additional instances will be conducted. Given that the lexicographic approach without slack, while yielding marginally better results on some metrics, incurs a significantly higher computational cost, it will be excluded from the next comparative analysis. Instead, we will focus on the lexicographic approach with slack and the scalar approach, as they offer a better balance between solution quality and efficiency, particularly in minimizing experience and competence violations within acceptable runtimes.

3.5.2 Results: 1 unit : 32 Nurses, 4 Weeks

Scalar Approach: $12Z_1 + 8Z_2 - 6Z_3 + 4Z_4 + 2Z_5$

Metric	**Lexicographic with slack**	**Scalar Approach**
Understaffing	**0.00**	**0.00**
Hours Deviation	78.25	78.25
Max Deviation	3.38	3.38
Overstaffing	60.00	60.00
Preference satisfaction	96.4%	96.4%
Experience violation	0	0
Competence violation	0	0
Run time (s)	330.55	325.80

Table 3.4: Comparison of multi-objective approaches on an instance with 32 nurses and 1 unit

3.5.3 Results: 2 units : 65 Nurses, 4 Weeks

Scalar Approach: $12Z_1 + 8Z_2 - 6Z_3 + 4Z_4 + 2Z_5$

Metric	**Lexicographic with slack**	**Scalar Approach**
Understaffing	**14.00**	**0.00**
Hours Deviation	**15.5**	22.85
Max Deviation	**0.38**	1.00
Overstaffing	41.00	**42.00**
Preference satisfaction	99.3%	**99.5%**
Experience violation	0	0
Competence violation	0	0
Run time (s)	535.39	**242.80**

Table 3.5: Comparison of multi-objective approaches on an instance with 65 nurses and 2 units

3.5.4 Results: 3 units 104 Nurses, 4 Weeks

The workforce consists of 104 nurses distributed across three sections as shown in Table 3.7.

Section	**Number of Nurses**
Section 1	32
Section 2	33
Section 3	39
Total	**104**

Table 3.6: Distribution of Nurses by Hospital Section

Scalar Approach : $12Z_1 + 8Z_2 - 6Z_3 + 4Z_4 + 2Z_5$

Metric	**Lexicographic (with slack)**	**Scalar Approach**
Understaffing	42.00	**0.00**
Hours Deviation	270.10	**253.65**
Max Deviation	**2.25**	4.25
Overstaffing	**232**	180
Preference satisfaction	**99.1**%	99.0%
Experience violation	**0**	2
Competence violation	**0**	2
Run time (s)	6147	**3600**

Table 3.7: Comparison of multi-objective approaches on an instance with 104 nurses and 3 units

The results obtained on different instance sizes highlight the trade-offs between the lexicographic approach with slack and the scalar multi-objective formulation. On a small instance with a single unit and 32 nurses (Table 3.4), both methods yield identical outcomes in terms of coverage, preference satisfaction, and constraint compliance, with a slightly shorter runtime for the scalar approach.

For the medium-sized instance with two units and 65 nurses (Table 3.5), the scalar approach completely eliminates understaffing, whereas the lexicographic method records 14 uncovered shifts. This improvement comes with a moderate increase in total extra hours, but the execution time is more than halved, showing a significant gain in computational efficiency.

The largest instance involves three units and 104 nurses distributed in sections, as shown in Table 3.7. In this scenario, the scalar approach again achieves perfect coverage but incurs minor violations in experience and skill constraints. However, it reduces total overtime and significantly improves runtime performance, decreasing the solution time from 6147 seconds to 3600 seconds.

3.6 Comparison of Baseline and Data-Driven Scheduling Models

In this section, we evaluate the effectiveness of the proposed heuristic in managing absences based on the initial schedules generated by two different models: the baseline version and the enhanced data-driven version introduced in Chapter 2. The objective is to measure how well the heuristic performs in covering daily absences over a realistic planning horizon and to assess the overall quality of the resulting schedules.

The evaluation is carried out using four real-world instances that combine two staffing scales and two absence contexts. A small-scale case with 32 nurses and a large-scale case with 104 nurses are each tested on two distinct months: one from 2018, representing a period with moderate absenteeism, and one during the COVID-19 crisis, characterized by high and volatile absences. This setup allows for a consistent comparison of performance across different planning scales and stress conditions, thereby assessing both the robustness and scalability of the proposed approach.

Absence Coverage Performance

The performance of the heuristic in covering daily absences was evaluated over a 28-day rolling horizon for both the baseline and data-driven initial schedules.

Simulation Results for 32 Nurses, 1 Unit – Moderate Absence Period (July 2022)

Metric / KPI	**Baseline Model**	**Data-Driven Model**
Absence Coverage		
Total Simulated Absences	47	47
Total Covered Absences	11	**14**
Coverage Rate (%)	23.4%	**29.8%**
Scheduling Quality		
Understaffing (nurse-shifts)	0.00	0.00
Total Hours Deviation (hours)	66.25	78.25
Max Individual Deviation (hours)	2.38	3.38
Experience Violations	0	0
Competence Violations	0	0
Preference Satisfaction (%)	96.1	**96.4**
Run Time (seconds)	**112.98**	302.58

Table 3.8: Global Summary of Simulation Results and Scheduling KPIs (July 2022, 32 Nurses)

Simulation Results for 32 Nurses, 1 Unit – High Absence Period (COVID-19)

Metric / KPI	**Baseline Model**	**Data-Driven Model**
Absence Coverage		
Total Simulated Absences	126	126
Total Covered Absences	24	**32**
Coverage Rate (%)	19.0%	**25.4%**
Scheduling Quality		
Understaffing (nurse-shifts)	0.00	0.00
Total Hours Deviation (hours)	**66.25**	91.50
Max Individual Deviation (hours)	**2.38**	4.25
Experience Violations	0	0
Competence Violations	0	0
Preference Satisfaction (%)	96.1%	**97.7%**
Run Time (seconds)	**122.81**	406.75

Table 3.9: Global Summary of Simulation Results and Scheduling KPIs (COVID-19 Period, 32 Nurses)

Comparison between baseline and data-driven models across both moderate and high absence periods reveals consistent improvements in absence coverage when using the data-driven approach. In the moderate absence scenario (July 2022), the data-driven model achieved a coverage rate of 29.8%, compared to 23.4% for the baseline. This performance gap widened under the more challenging conditions of the COVID-19 peak, where the data-driven model covered 25.4% of absences, versus only 19.0% for the baseline. These results suggest that the data-driven model is more effective in anticipating and structurally absorbing potential absences.

In terms of scheduling quality, both models maintained zero understaffing and no violations of experience or competence constraints. Although the total hours deviation is slightly higher for the data-driven model, this metric aggregates deviations across all nurses and is therefore less informative on its own. What is more relevant is the maximum individual hours deviation, which indicates the highest workload imbalance among staff. This value increases from 2.38 hours in the baseline to 3.38 hours in the data-driven model during the moderate absence period, and to 4.25 hours during the high absence period. This suggests a slightly higher individual burden in the data-driven scenario, likely as a trade-off for improved absence coverage. Nevertheless, preference satisfaction remains high in both cases, with a slight advantage for the data-driven model.

Simulation Results for 104 Nurses, 3 Units – Moderate Absence Period(January 2018)

Metric / KPI	**Baseline Model**	**Data-Driven Model**
Absence Coverage		
Total Simulated Absences	278	278
Total Covered Absences	82	**115**
Coverage Rate (%)	29.5%	**41.4%**
Scheduling Quality		
Understaffing (nurse-shifts)	0.00	0.00
Total Hours Deviation (hours)	**236.40**	286.15
Max Individual Deviation (hours)	4.25	4.25
Experience Violations	0	0
Competence Violations	0	0
Preference Satisfaction (%)	98.6%	**98.8%**

Table 3.10: Global Summary of Simulation Results and Scheduling KPIs (January 2018, 104 Nurses)

Simulation Results for 104 Nurses, 3 Units – High Absence Period (COVID-19)

Metric / KPI	**Baseline Model**	**Data-Driven Model**
Absence Coverage		
Total Simulated Absences	336	336
Total Covered Absences	78	**124**
Coverage Rate (%)	23.2%	**36.9%**
Scheduling Quality		
Understaffing (nurse-shifts)	0.00	0.00
Total Hours Deviation (hours)	**236.65**	286.15
Max Individual Deviation (hours)	4.25	4.25
Experience Violations	0	0
Competence Violations	0	0
Preference Satisfaction (%)	98.6%	**98.8%**

Table 3.11: Global Summary of Simulation Results and Scheduling KPIs (COVID-19 Period, 104 Nurses)

The results observed on the larger-scale test case, involving 104 nurses and three care units, confirm the trends highlighted in the smaller-scale simulations. The data-driven model consistently outperforms the baseline in terms of absence coverage, regardless of the level of absenteeism. During the moderate absence period (January 2018), the data-driven model covers 41.4% of the absences compared to 29.5% for the baseline. This improvement is even more pronounced during the COVID-19 period, where the data-driven model achieves a coverage rate of 36.9%, while the baseline reaches only 23.2%.

Regarding scheduling quality, both models ensure zero understaffing and full compliance with experience and competence constraints, even under high pressure. As in previous scenarios, the total hours deviation is slightly higher for the data-driven model. However, the maximum individual deviation remains identical at 4.25 hours for both models, indicating that no additional burden is placed on individual staff members despite the higher global coverage.

Preference satisfaction remains excellent in both configurations and slightly favors the data-driven model. One noticeable difference lies in the optimization gap: while the baseline model converges to tighter bounds (3.07%–3.41%), the data-driven model exhibits a wider gap (around 14%). This suggests that while the data-driven solution achieves better operational performance, it may require more advanced optimization strategies or tuning to reduce solution uncertainty in larger instances.

3.7 Data Analysis

3.7.1 Exploratory Absence Analysis

As a first step, we conducted an in-depth exploratory analysis of historical staffing records. The aim was to better understand the dynamics of absenteeism by comparing the number of nurses scheduled (*planned*) to the number who actually reported to work (*actual*) across three years of data, for each section and shift.(This evaluation was conducted using our large-scale dataset of 104 nurses and 3 units)

This analysis was structured in three stages:

1. **Boxplot visualizations** were used to capture the distribution of absences on weekdays and shifts, highlighting variability, outliers and central trends within each section.
2. We then examined the **weekly distribution of absences in percentage terms**, separately for each shift. This allowed us to identify which days concentrate the highest share of absences in a typical week without comparing shifts directly.
3. Finally, we introduced a **shift-day level metric** to support planning decisions: the *criticality score*, which combines both the intensity and frequency of absences for each (weekday, shift, section) combination.This score is the one used in the objective function of the data-driven scheduling model, allowing us to assign differentiated weights across (weekday,shift) combinations. It provides a quantitative basis for comparing staffing pressure between time slots and ensures that the model prioritizes the most critical periods when reallocating resources.

3.7.1.1 Boxplot visualizations

We begin this analytical sequence with the first component: the exploration of daily absenteeism patterns using boxplot visualizations. This step aims to assess the variability and distribution of absences across shifts and weekdays, rather than focusing solely on average rates. By observing the spread, concentration, and outliers of daily absences, we gain a clearer picture of which shifts exhibit the greatest instability and which days are more prone to extreme absenteeism.

The analysis was conducted separately for each section to preserve the specific structural and operational characteristics of each unit.

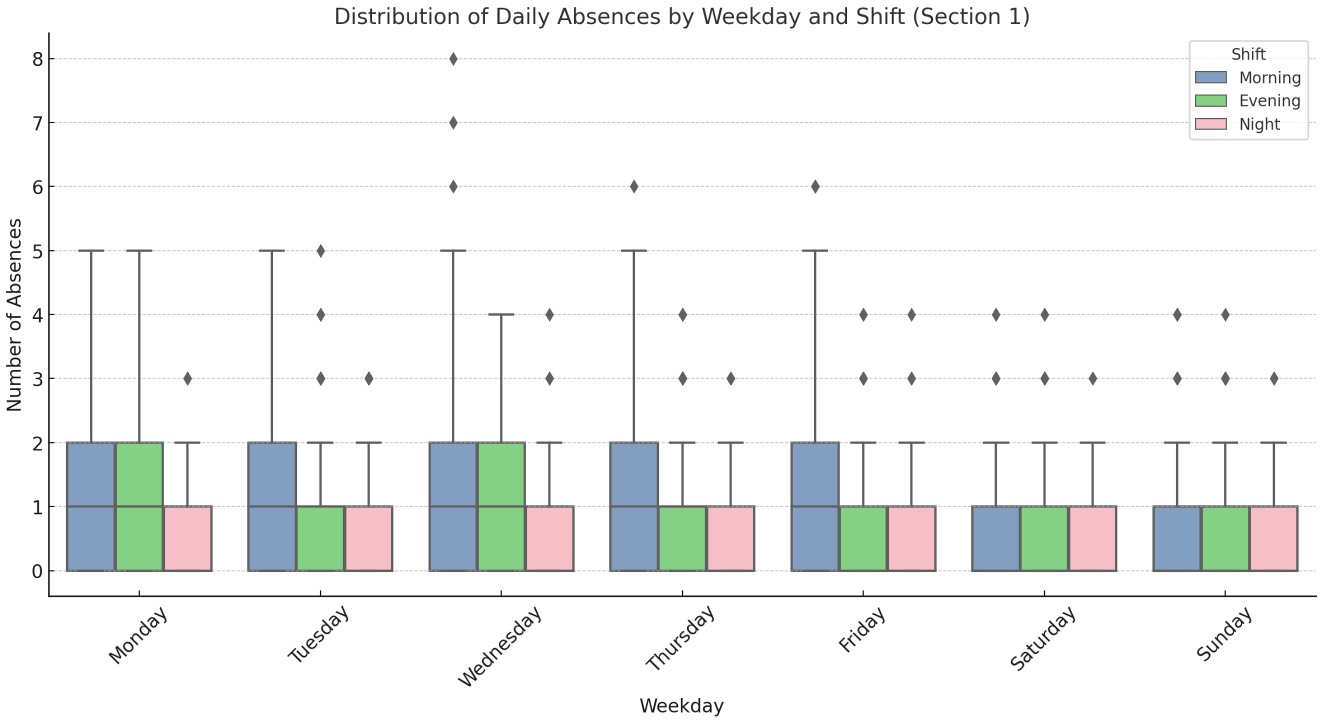

Figure 3.5: The boxplots for Section 1

The boxplot for Section 1 shows clear differences between the three shifts. Morning shifts have the highest variation in absences. On weekdays, especially from Tuesday to Friday, some days had very high numbers of absences. The spread is wide, which means that the number of absentees changes a lot from one day to another.

In evening shifts, the number of absences is usually lower and more stable. Most days have 0 to 2 absences, and large values are rare. There is still some variation, especially at the start of the week.

Night shifts are the most stable. Absences are low on most days, often just 0 or 1. There are very few extreme cases. Across all shifts, weekends show a noticeable reduction in both the median number of absences and their variability. Weekdays, on the other hand, are marked by increased fluctuations, especially during the morning shift. These trends are consistent across the entire historical window.

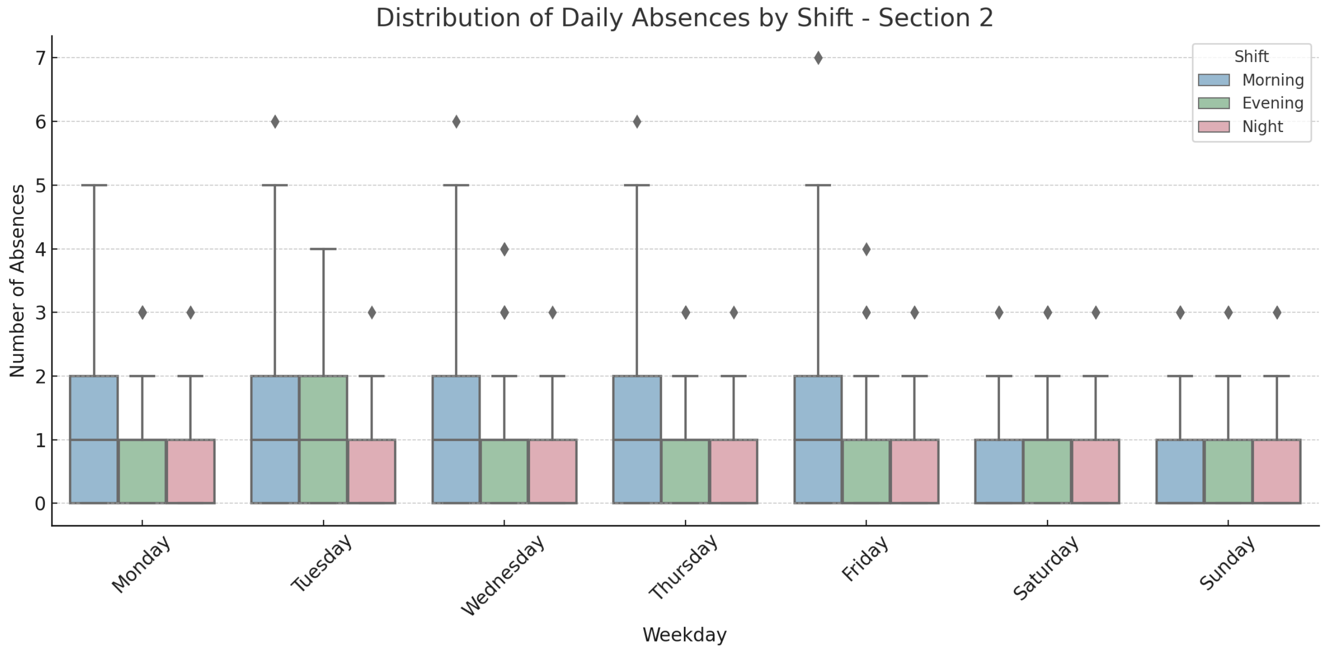

Figure 3.6: The boxplots for Section 2

The absenteeism profile in Section 2 is slightly more balanced compared to Section 1, but some differences still appear between shifts and weekdays. The morning shifts show a clear concentration of higher absences, especially on weekdays. The spread is moderate to wide, indicating that some days have very few absences and others have higher, less predictable numbers. Tuesday, Wednesday, and Friday show larger spreads, suggesting more staffing instability.

In the evening shifts, the variation in absences is smaller. Most values are centered on low counts, though some days show moderate peaks, particularly midweek. The pattern seems to be more stable than in the mornings.

Night shifts remain the most stable and predictable. The absences here are typically very low, often 0 or 1, with only occasional deviations. The consistency suggests a lower risk of absence during night shifts.

Across all shifts, weekends again show fewer absences, with smaller ranges and fewer outliers consistent with reduced demand or different working dynamics during weekends.

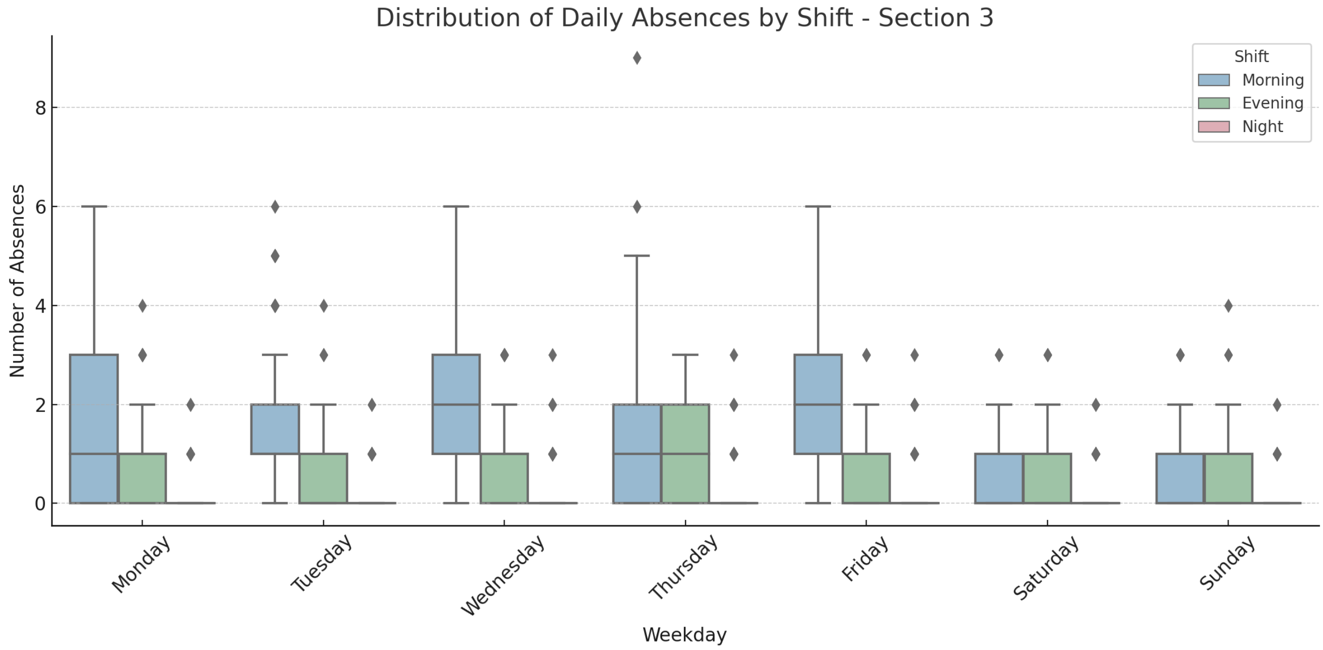

Figure 3.7: The boxplots for Section 3

In Section 3, absenteeism patterns are more pronounced than in the other sections, particularly for morning shifts, which show both high central tendency and wide dispersion. The upper quartiles and several outliers on Wednesday, Thursday, and especially Friday confirm a recurring problem in morning coverage on those days.

Evening shifts show moderate absenteeism with relatively tight distributions during the week. Although the variation is not as high as that for mornings, certain weekdays, such as Thursday, still show extended ranges, indicating potential planning challenges.

Night shifts continue to display the most stable attendance across the week. Absences are low and consistent, with almost no outliers, reinforcing the idea that night coverage is generally more reliable.

Once again, weekend shifts show the smallest absenteeism across all three periods, with tight boxes and fewer extreme values. This suggests lighter staffing needs or possibly stronger attendance culture on weekends in this section.

Weekly Distribution of Absences

To analyze the concentration of absences throughout the week, we computed the share of total absences occurring on each weekday, separately, for each shift. This allows for intra-shift comparisons while preserving temporal patterns. The percentage of absences for a given weekday d and shift s is calculated as follows:

$$\text{AbsenceShare}_{d,s} = \frac{\sum_{k=0}^{K} k \cdot N_{d,s}^{(k)}}{\sum_{d'} \sum_{k=0}^{K} k \cdot N_{d',s}^{(k)}} \times 100 \tag{3.1}$$

where:

- k is the number of absences observed on a given day,

- $N_{d,s}^{(k)}$ is the number of days with exactly k absences on weekday d and shift s,
- the numerator gives the total number of absences accumulated on weekday d,
- the denominator gives the total number of absences across all weekdays for shift s.

This metric highlights which weekdays contribute most to overall absenteeism in each shift, without comparing shifts directly.

The figures below illustrate this distribution for each section.

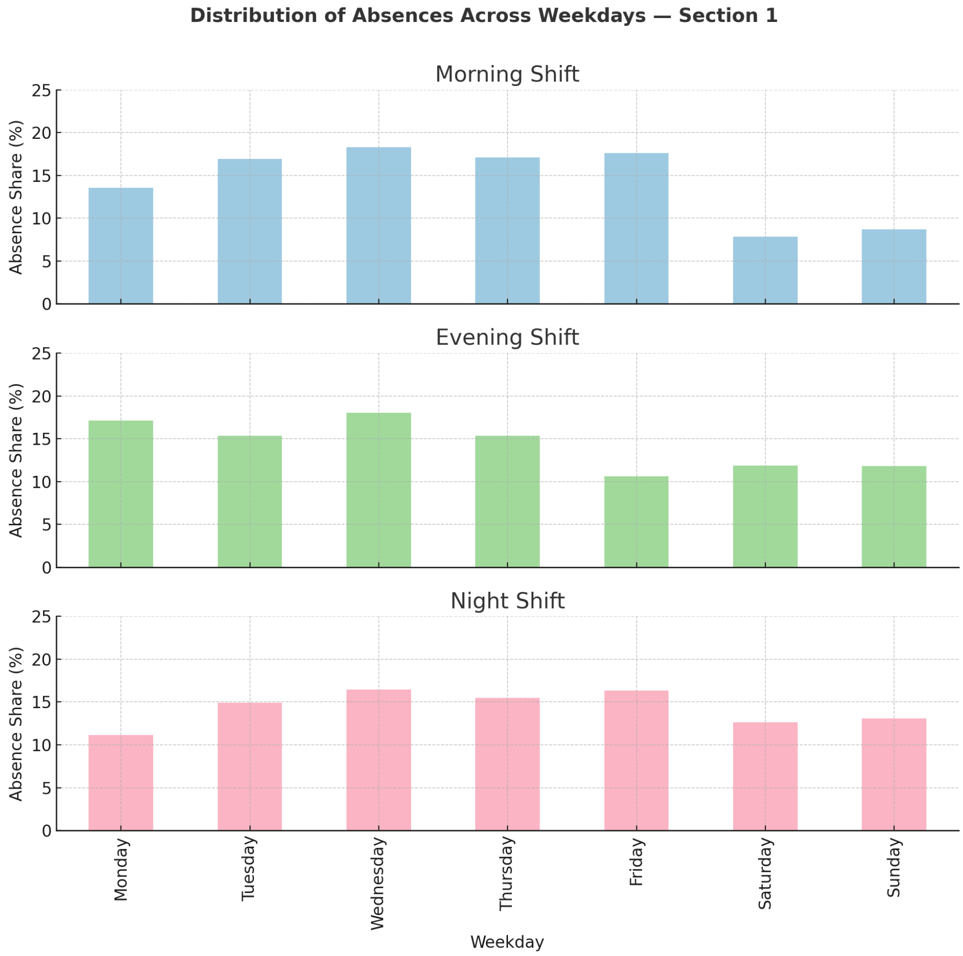

Figure 3.8: Distribution of Daily Absences by Weekday and Shift (Section 1)

In Section 1, the **morning shift** shows a fairly balanced distribution of absences across the core weekdays, with Tuesday through Friday each accounting for roughly 17% to 19% of total absences. Monday is slightly lower (~13%), while Saturday and Sunday clearly show reduced absenteeism, each contributing under 10%. This confirms that most morning shift absences are concentrated during the typical workweek.

The **evening shift** displays a more uneven profile. Monday and Wednesday show the highest proportions (~17–18%), suggesting increased absenteeism early to midweek. Friday, on the other hand, accounts for the lowest share, around 10%, with weekend values slightly higher (around 12%), resulting in a modest U-shaped trend across the week.

For the **night shift**, absences are more evenly spread across all days, but a slight increase is visible midweek. Wednesday and Friday reach the highest values (~16%), while Monday is the lowest at approximately 11%. Unlike the morning shift, there is no clear drop in the weekend shift in the night shift, indicating a more stable distribution throughout the week.

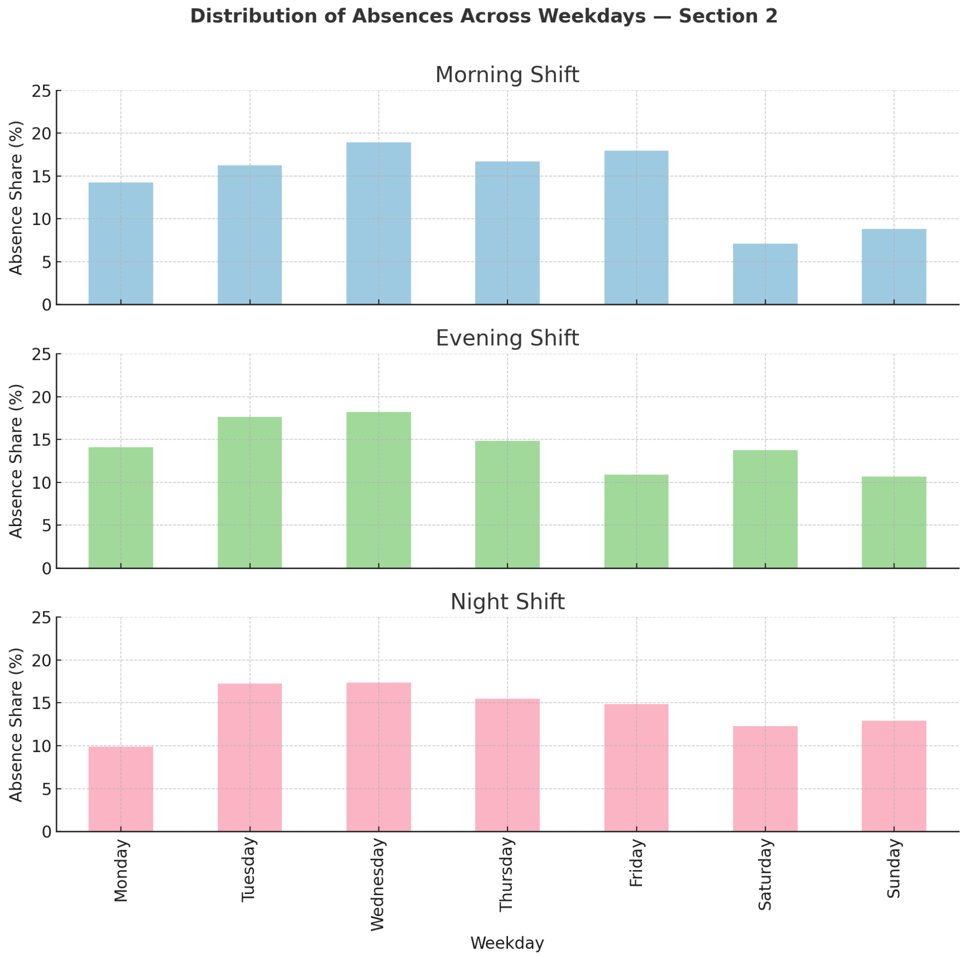

Figure 3.9: Distribution of Daily Absences by Weekday and Shift (Section 2)

The **morning shift** shows a clear concentration of absences on the weekday, with Wednesday and Friday recording the highest shares (both close to 19%). Tuesday and Thursday also contribute significantly, maintaining a fairly elevated level of absenteeism across the central workweek. In contrast, Saturday and Sunday are visible lower, accounting for around 7% and 9%, respectively, confirming a drop in absences during the weekend.

In the **evening shift**, the distribution is more varied. Tuesday and Wednesday stand out with the highest proportions, each reaching nearly 18% of the weekly evening absences. Monday follows at a moderate level, while Friday, Saturday, and Sunday show reduced shares, especially Sunday which drops just above 10%. This indicates a concentration of evening shift absences earlier in the week.

For the **night shift**, the distribution is more polarized. Tuesday and Wednesday dominate the week, each accounting for more than 17% of night shift absences. Monday is the lowest day (around 10%), and the weekend remains relatively low and stable (about 12–13%). This confirms a stronger mid-week absenteeism load for night shifts in Section 2.

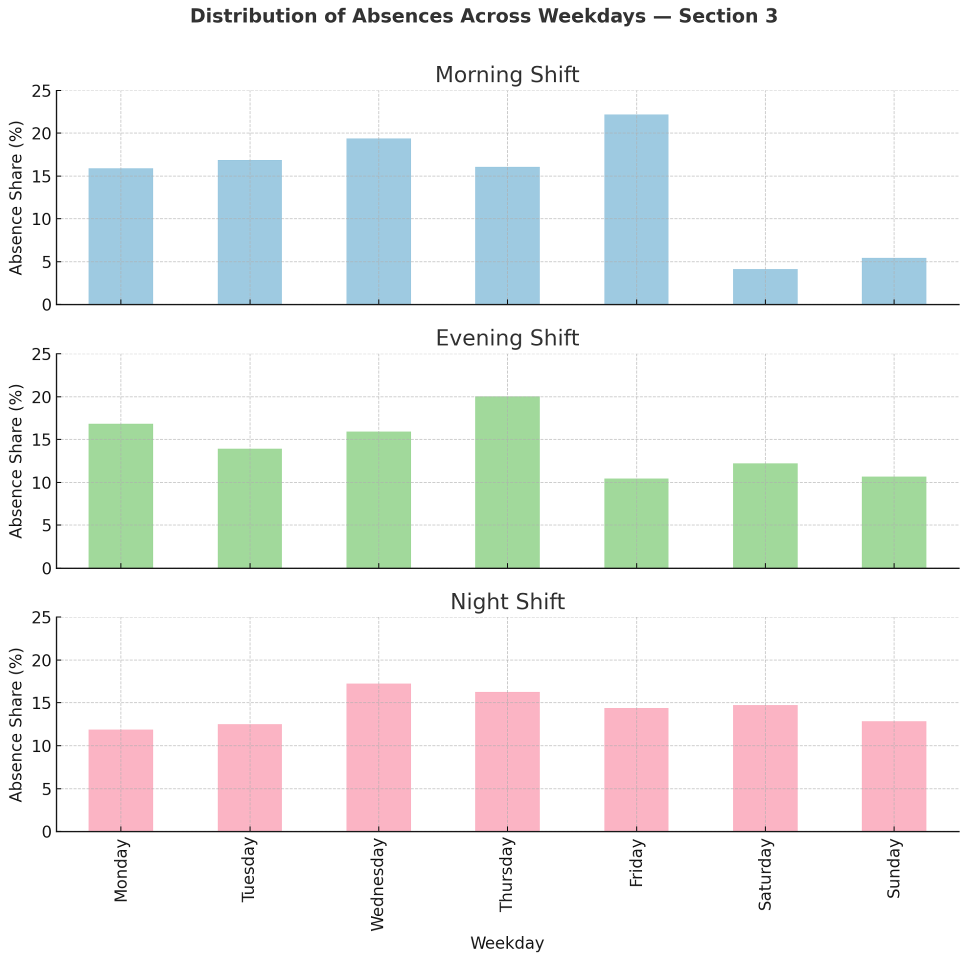

Figure 3.10: Distribution of Daily Absences by Weekday and Shift (Section 3)

In Section 3, the **morning shift** shows the most pronounced weekly concentration of absences among all three sections. Friday alone accounts for over 22% of total morning absences, followed by Wednesday (19%) and Tuesday (17%). In contrast, Saturday and Sunday register very low proportions, each contributing around 4% to 6%, highlighting a strong weekday dominance in absenteeism.

For the **evening shift**, the distribution is slightly more balanced but still variable. Thursday emerges as the peak day, concentrating 20% of evening shift absences. Monday and Wednesday also show elevated shares (~16–17%). In contrast, Friday, Saturday, and Sunday remain the least affected, all below or around 12%, suggesting a reduced evening absenteeism towards the end of the week.

The **night shift** displays a more uniform distribution of absences throughout the week. The shares range from 12% to 17%, with no day defining with a sharp difference. Wednesday and Thursday show slightly higher values, while Monday and Sunday are at the lower end. This regularity reinforces the overall stability of night shift attendance in Section 3.

3.7.1.2 criticality score

To complement the previous analyses on absenteeism distributions, we now introduce a composite indicator that captures both the intensity and the recurrence of absenteeism: the **Criticality Score** C_{ads}. Unlike raw frequency or percentage-based metrics, this score takes into account the actual number of absences on each day and how often each level of absence occurs. It provides a more nuanced understanding of which weekday–shift combinations truly concentrate high-risk staffing situations.

The following table summarizes the values of this score across all three sections and shifts, offering a comparative view of the most critical periods throughout the week.

- **Criticality Score** C_{ads}:

$$C_{ads} = \frac{\sum_a a \times D_{ads}^{(a)}}{N_d^{\text{total}}} \tag{3.2}$$

where:

- a: A specific number of absences (e.g., 1, 2, 3, ...)
- $D_{ads}^{(a)}$: Number of days (among weekday d) where exactly a absences occurred in shift s, in the given section
- N_d^{total}: Total number of occurrences of weekday d in the dataset

More detailed results and a breakdown by section can be found in: 1.1 1.2 1.3

Weekday	Section 1			Section 2			Section 3		
	Morning	Evening	Night	Morning	Evening	Night	Morning	Evening	Night
Monday	1.084	0.962	0.395	1.038	0.720	0.299	1.536	0.816	0.146
Tuesday	1.352	0.862	0.529	1.184	0.900	0.521	1.628	0.674	0.153
Wednesday	**1.464**	**1.015**	**0.582**	**1.379**	**0.931**	**0.525**	1.874	0.770	**0.211**
Thursday	1.368	0.862	0.548	1.215	0.759	0.467	1.559	**0.969**	0.199
Friday	1.410	0.598	0.579	1.307	0.556	0.448	**2.146**	0.506	0.176
Saturday	0.628	0.667	0.448	0.517	0.701	0.372	0.402	0.590	0.180
Sunday	0.700	0.665	0.465	0.646	0.546	0.392	0.531	0.519	0.158

Table 3.12: Weekly criticality comparison by shift

The criticality table highlights several consistent patterns across the three hospital sections. Morning shifts systematically exhibit the highest criticality scores in all sections, particularly on Wednesdays and Fridays. For example, in Section 3, the Friday morning shift reaches a peak of 2.146, indicating both a high number of absences and frequent recurrence.

In Section 1, Wednesday mornings also stand out (1.464), along with Friday (1.410), confirming that mid-to-late week morning shifts are generally the most strained. Section 2 displays a similar pattern, although with slightly lower criticality levels overall, particularly in the evening and night shifts.

Evening shifts show more moderate criticality scores. In Section 1 and 2, Wednesday evenings are the most problematic, while in Section 3, the peak shifts slightly to Thursday evening (0.969). The contrast with Friday evenings, especially in Sections 1 and 2, is notable and suggests lower planning risk toward the weekend.

As expected, night shifts consistently report the lowest criticality across all sections. However, the scores are not negligible. Wednesday and Friday nights still show peaks (e.g., 0.582 and 0.579 in Section 1), suggesting that while rarer, night absences can still present operational challenges on certain days.

In all sections, weekends (Saturday and Sunday) display uniformly lower criticality values for all shifts, reinforcing the patterns seen in earlier boxplots and bar charts: staffing pressure is markedly reduced on weekends.

3.7.2 Predictive Analysis of Staff Absenteeism

The objective of this analysis is to predict the number of absentees per shift on a daily basis. Since the target variable is a count (the number of nurses absent), this constitutes a **regression task**, not a classification problem. We are not simply interested in determining whether an absence occurs, but rather in estimating its magnitude. This distinction is fundamental for selecting appropriate predictive models and evaluation metrics.

To develop a robust and well-adapted model, we performed a thorough diagnostic of the statistical structure of the data. The aim was to detect any distributional patterns such as zero-inflation, over-dispersion, or skewness that would invalidate the assumptions of classical regression models (e.g linear or Poisson) and instead motivate the adoption of more flexible modeling strategies like the Hurdle model.

Three empirical patterns were identified:

1. Excess Zeros (Zero-Inflation)

A substantial number of days exhibited no absences at all, particularly during night shifts. In some configurations, the proportion of zeros exceeded 80%, as reported in Table **??**. This zero-inflation violates the assumptions of standard count models like Poisson or Negative Binomial, which typically underestimate the frequency of zeros.

Section	Shift	Zero Proportion
1	D	38.6%
1	A	46.2%
1	N	61.8%
2	D	41.3%
2	A	49.2%
2	N	64.0%
3	D	34.8%
3	A	50.8%
3	N	84.1%

Table 3.13: Proportion of zero absences per shift and section

2. Overdispersion (Variance > Mean)

We formally assessed whether the data showed an overdispersion where the variance exceeds the mean, which contradicts the core assumption of the Poisson distribution ($\mathrm{Var}(Y) = \mathrm{E}(Y)$). Following [68], we compute the Pearson dispersion index $\tilde{\phi}$:

$$\tilde{\phi} = \frac{X^2}{n-p}, \quad \text{where} \quad X^2 = \sum \frac{(y_i - \bar{y})^2}{\bar{y}}$$

An index $\tilde{\phi} > 1$ indicates overdispersion. The results shown in Table **??** confirm mild to moderate overdispersion in most combinations of shift sections.

Section	**Shift**	**Mean**	**Variance**	$\tilde{\phi}$ (Pearson)
1	D	1.14	1.51	**1.32**
1	A	0.80	0.85	**1.06**
1	N	0.51	0.54	**1.06**
2	D	1.04	1.38	**1.33**
2	A	0.73	0.75	**1.02**
2	N	0.43	0.40	0.94
3	D	1.38	1.93	**1.39**
3	A	0.69	0.68	0.98
3	N	0.17	0.18	**1.02**

Table 3.14: Dispersion analysis based on mean, variance, and Pearson index

3. Skewed and Sparse Count Distributions

Figure 3.11 illustrates that all shift distributions are right-skewed, with a spike at zero and a long positive tail. This pattern reinforces the need for a model that separates zero generation from count intensity.

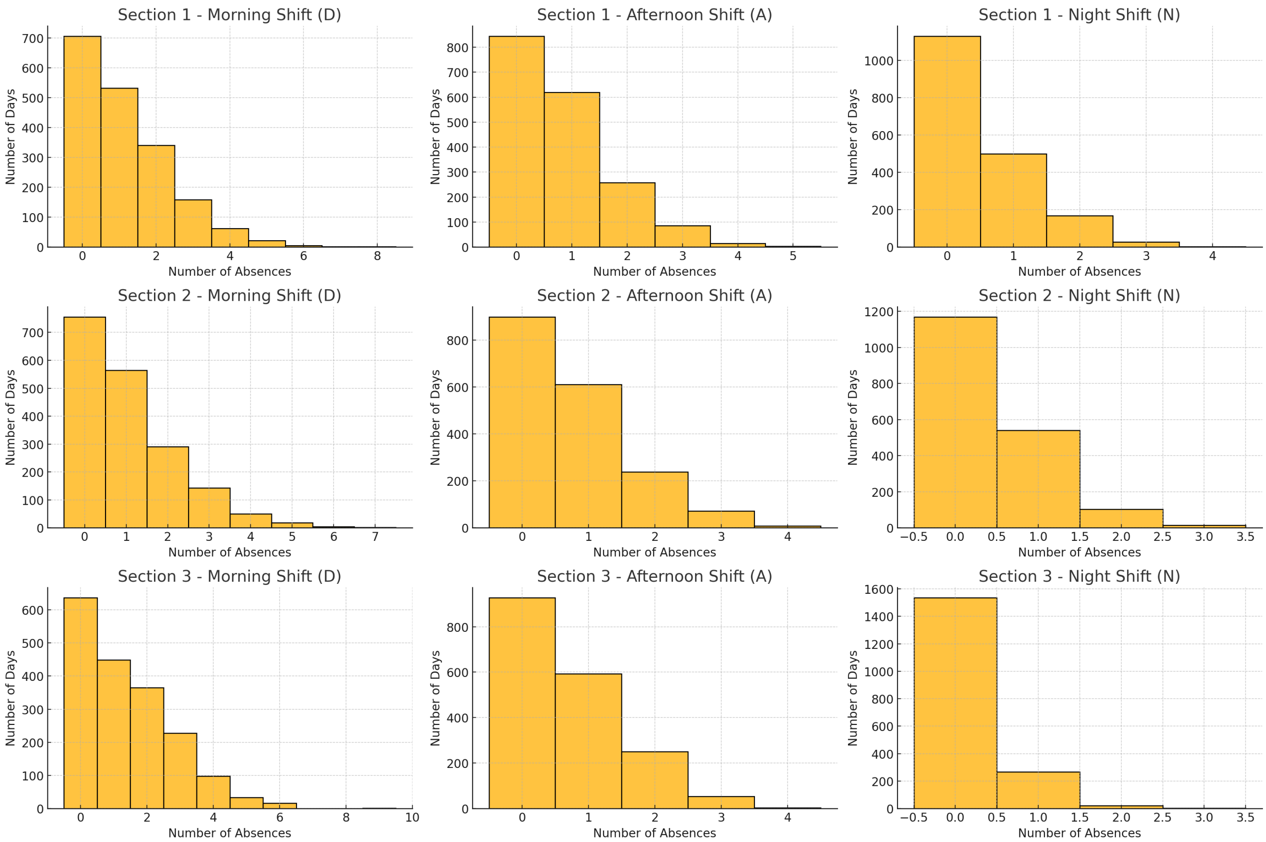

Figure 3.11: Histograms of Absences – Distributional Shape

Modeling Strategy: The Hurdle Model

These three findings: excess zeros, overdispersion, and skewness jointly motivate the use of a two-part count model. We implemented a **Hurdle model**, which separately models:

- the *occurrence* of absences using logistic regression;
- the *intensity* of absences (conditional on being positive) using truncated count models.

Implementation and Results

Step 1: Predicting Occurrence (Logistic Regression) We trained a logistic regression to predict whether absences would occur ($y > 0$), using calendar and workload-related variables. Performance was evaluated using McFadden's pseudo-R^2 and the Area Under the Curve (AUC).

McFadden's pseudo-R^2 measures how much better the model fits the data compared to a model without predictors. Unlike R^2 used in linear regression, its values are usually much lower. According to Hu et al. [69], values between **0.1 and 0.3** already suggest a model with a good fit. Even when predictors have a strong effect, pseudo-R^2 values often stay below 0.2, which is normal in logistic regression.

The AUC indicates how well the model distinguishes between days with and without absences. An AUC of 0.5 means random guessing, while values closer to 1.0 show that the model ranks positive cases (days with absence) higher than negative ones. According to Google Developers [70], AUC values between **0.7 and 0.8** are considered acceptable, between **0.8 and 0.9** are good, and above **0.9** are excellent.

Section	Shift	Pseudo-R^2 (McFadden)	AUC
1	D	**0.121**	0.731
1	A	0.094	0.704
1	N	**0.448**	0.886
2	D	**0.182**	0.777
2	A	**0.211**	0.798
2	N	**0.311**	0.826
3	D	**0.190**	0.788
3	A	**0.205**	0.797
3	N	**0.204**	0.786

Table 3.15: Performance of logistic models for predicting absence occurrence

Step 2: Predicting Intensity (Count Regression on Positive Cases) For cases with absences ($y > 0$), we compared four models:

- Hurdle model (conditional part : Linear Regression, poisson, negative binomial)
- Linear Regression
- Random Forest
- XGBoost

Target Variables and Feature Set : We built separate regression models for each shift: morning (`abs_D`), afternoon (`abs_A`), and night (`abs_N`). These three count targets were predicted using the following features:

- `Days`: day of the week (0–6).
- `is_weekend`, `is_holiday`, `month`: calendar variables.
- `n_sejours`: number of inpatients.
- `D_plan`, `A_plan`, `N_plan`: number of planned staff per shift.

All models were trained and evaluated using an 80/20 train-test split, ensuring consistent evaluation across configurations.

Performance was evaluated using MAE = Mean Absolute Error, RMSE = Root Mean Squared Error.

RF = Random Forest, LR = Linear Regression, lin = linear, pois = poisson, nb = negative binomial.

Shift	Hurdle (lin)	XGBoost	RF	LR	Poisson	Hurdle (pois)
MAE						
abs_D	0.798	0.798	0.818	**0.794**	0.824	0.804
abs_A	0.663	0.684	0.704	0.678	0.689	**0.661**
abs_N	0.321	0.324	0.325	0.397	0.537	**0.315**
Shift	**Hurdle (lin)**	**Hurdle (nb)**	**XGBoost**	**RF**	**LR**	**Poisson**
RMSE						
abs_D	**0.804**	1.005	1.021	1.034	1.002	1.011
abs_A	**0.661**	0.845	0.861	0.889	0.856	0.868
abs_N	**0.320**	0.455	0.458	0.472	0.496	0.630

Table 3.16: Comparaison des performances prédictives sur la section 1 (MAE et RMSE par shift)

Shift	Hurdle (lin)	XGBoost	RF	LR	Poisson	Hurdle (pois)
MAE						
abs_D	0.750	**0.748**	0.765	0.757	0.795	0.756
abs_A	0.546	**0.529**	0.533	0.572	0.626	0.546
abs_N	0.339	0.346	0.360	0.391	0.507	**0.338**
Shift	**Hurdle (lin)**	**Hurdle (nb)**	**XGBoost**	**RF**	**LR**	**Poisson**
RMSE						
abs_D	**0.757**	0.955	0.967	0.985	0.959	0.986
abs_A	**0.549**	0.686	0.671	0.698	0.707	0.750
abs_N	**0.340**	0.450	0.463	0.481	0.486	0.576

Table 3.17: Comparaison des performances prédictives sur la section 2 (MAE et RMSE par shift)

Shift	Hurdle (lin)	XGBoost	RF	LR	Poisson	Hurdle (pois)
MAE						
abs_D	**0.897**	0.899	0.931	0.901	0.919	0.903
abs_A	**0.532**	0.544	0.577	0.542	0.611	**0.532**
abs_N	**0.244**	0.256	0.259	0.277	0.309	**0.244**
Shift	**Hurdle (lin)**	**Hurdle (nb)**	**XGBoost**	**RF**	**LR**	**Poisson**
RMSE						
abs_D	**0.902**	1.144	1.153	1.189	1.149	1.158
abs_A	**0.533**	0.683	0.701	0.734	0.695	0.733
abs_N	**0.245**	0.396	0.398	0.412	0.410	0.460

Table 3.18: Comparaison des performances prédictives sur la section 3 (MAE et RMSE par shift)

The analysis of predictive performance, measured by MAE and RMSE across all sections and shifts, clearly highlights the superiority of the **linear Hurdle model**. This model consistently achieves the best or joint-best results across all shifts and all three sections. In particular, it records the lowest root mean squared errors (RMSE) for **each shift in every section**, demonstrating its robustness in predicting absences, even for less frequent shifts, such as night shifts. While other models such as XGBoost or the Poisson Hurdle model occasionally perform well on specific shifts or sections in terms of MAE, none manage to outperform the linear Hurdle model overall. This joint dominance in both average accuracy and error stability justifies its selection as the **benchmark model for the subsequent analyses**.

In summary, the Hurdle model is both statistically justified and empirically competitive across shifts and sections, especially for night shifts where standard models struggle with zero-inflated and skewed distributions.

Comparative Evaluation of Scheduling Models Using Hurdle-Based Absence Predictions

After comparing several approaches for simulating absences, predictive analysis demonstrated the superior performance of the Hurdle model, particularly due to its ability to model both the occurrence and the intensity of absences separately. Based on this finding, we now use the absences predicted by the Hurdle model as a fixed input to evaluate and compare the performance of the two scheduling approaches: the baseline model and the data-driven model. The following table 3.19 summarizes the key performance indicators obtained over a 28-day horizon, with a total of 193 simulated absences.

Metric / KPI	Data-Driven Model	Baseline Model
Absence Coverage		
Total Simulated Absences	193	193
Total Covered Absences	**84**	53
Coverage Rate (%)	**43.5%**	27.5%
Scheduling Quality		
Understaffing (nurse-shifts)	0.00	0.00
Total Hours Deviation (hours)	280.15	**249.65**
Max Individual Deviation (hours)	4.25	4.25
Experience Violations	**0**	2
Competence Violations	**0**	2
Preference Satisfaction (%)	98.8%	**99.1%**

Table 3.19: Simulation Summary with Absences Predicted by Hurdle Model (28 Days, 193 Total Absences)

The results in Table 3.19 reinforce the conclusions drawn from previous scenarios: the data-driven model significantly outperforms the baseline in terms of absence coverage, managing to absorb **84 out of 193 predicted absences**, compared to only **53** for the baseline. This translates to a **coverage rate of 43.5%**, versus **27.5%**, confirming that proactive allocation of buffer capacity toward high-risk shifts, guided by historical absence patterns, effectively improves the resilience of the system.

Both models maintain complete compliance with the staffing requirement, achieving **zero understaffing** across the entire horizon. However, this gain in coverage comes with a moderate increase in *total hours deviation* (+30.5 hours), a logical consequence of the greater flexibility introduced in the data-driven schedule. Despite this, the *maximum workload imbalance per nurse remains identical* in both cases (4.25 hours), showing that the additional adjustments are spread across the team rather than concentrated on a few individuals.

In addition, the data-driven model fully respects both **experience and competence requirements**, while the baseline solution violates each of these constraints twice. This demonstrates the added value of incorporating risk-aware allocation mechanisms that prioritize not only coverage, but also regulatory and professional standards of care delivery.

Finally, *preference satisfaction* remains high for both approaches, with a slight advantage for the baseline (99.1% vs. 98.8%). This marginal difference confirms that the improvements in coverage and constraint satisfaction achieved by the data-driven model do not come at the expense of individual nurse preferences.

In general, the integration of predictive modeling into the scheduling process enables more effective anticipation of disruptions, leading to a more robust and regulatory-compliant schedule without significantly compromising workload balance or staff satisfaction.

Conclusion

This chapter presented a comprehensive computational study designed to evaluate the performance of the proposed initial nurse scheduling model, enhanced with a data-driven component. Through a series of tests on instances of increasing size, we compared two multi-objective optimization approaches, lexicographic with slack and scalar weighting in terms of coverage, workload balance, preference satisfaction, and computational efficiency.

The introduction of a weighted objective function, guided by a criticality score derived from historical absence patterns, enabled the generation of schedules that are more resilient to disruptions. Absences were simulated using a Hurdle model, which is particularly well-suited for discrete, skewed, and zero-inflated data. These simulated absences were then integrated into a rolling horizon simulation framework, where a local reallocation heuristic based on an overflow unit was applied day by day to mitigate their impact.

The results showed that the schedules generated by the data-driven model led to significantly higher absence coverage up to 43.5% compared to 27.5% for the baseline model, while maintaining compliance with the skills, experience, workload and preference constraints.

In summary, this study confirms the value of incorporating predictive analytics into the initial scheduling phase to enhance system resilience in the face of staff absences, even in the absence of explicit real-time rescheduling mechanisms.

General Conclusion

This project tackled the nurse scheduling problem in hospital settings by focusing on improving the robustness of initial planning in the face of staff absenteeism. Rather than relying on reactive rescheduling, we proposed a proactive approach that integrates historical data, predictive modeling, and optimization techniques to generate more resilient baseline schedules.

We first formulated a multi-objective mixed-integer programming model that respects operational constraints such as workload balance, legal limits, and skill requirements, while also integrating nurse preferences. A criticality score computed from historical absence patterns was used to weigh time periods according to their risk level, allowing the model to better anticipate disruptions.

A statistical exploration of absenteeism data revealed significant temporal structures, such as weekday effects and shift-based patterns, as well as zero-inflation and overdispersion. These findings motivated the use of a Hurdle model to predict absences per shift and day. This two-part model offered improved predictive accuracy and was used to simulate realistic absenteeism scenarios.

We then evaluated the robustness of baseline schedules in a rolling horizon simulation framework. Each day, a simple yet effective heuristic based on a buffer assignment strategy (overflow unit) was used to manage unexpected absences. The results showed that schedules generated by the data-driven model significantly outperformed those of the baseline model, covering up to 43.5% of absences versus 27.5%, while maintaining feasibility and high satisfaction of preferences and regulatory constraints.

In essence, this work demonstrates that combining predictive analytics with anticipatory optimization can significantly strengthen nurse scheduling under uncertainty. By embedding historical absence patterns directly into the planning process, we were able to generate schedules that are both operationally feasible and inherently more resilient to disruption. The use of a criticality-weighted objective function, informed by data, allowed the model to focus resources where and when they are most needed. In parallel, the Hurdle model provided a realistic simulation of absences, enabling robust stress-testing of schedules. Altogether, this framework provides a practical foundation for building smarter, more adaptive workforce planning systems in healthcare environments.

Future Perspectives

Building on this work, several avenues can be explored to further improve performance and generalizability:

- **Integrate a rescheduling optimization model (MIP)** to handle absences that cannot be resolved by the heuristic. This would provide a more structured and optimal way of reallocating staff when simple reassignments fail.

- **Combine multiple coverage strategies**, such as overflow plus cross-unit reassignment, with prioritization rules based on criticality or multi-skilled staff. This hybrid logic could enhance flexibility without compromising care quality.

- **Shift to nurse-level absence prediction**, allowing for more precise anticipation of disruptions and more targeted overstaffing on critical shifts. This granularity could improve both predictive power and resource utilization.

Bibliography

[1] Bradley, E. H., et al. (2017). *Operations research for health care decision-making.* Global Health Journal.

[2] Evans, G. W. (1984). *An overview of techniques for solving multi-objective mathematical programs.* Management Science.

[3] Cohon, J. L., & Marks, D. H. (1975). *A review and evaluation of multiobjective programming techniques.* Water Resources Research.

[4] Goicoechea, A., Hansen, D. R., & Duckstein, L. (1987). *Multiobjective Decision Analysis with Engineering and Business Applications.* Wiley.

[5] Teleb, M. A., & Azadivar, F. (1994). *Interactive simulation-optimization for stochastic multiobjective problems.* Simulation.

[6] Barták, R., Salido, M. A., & Rossi, F. (2008). *Constraint satisfaction techniques in planning and scheduling.* Journal of Intelligent Manufacturing, 21(1), 5–15.

[7] Ouelhadj, D., & Petrovic, S. (2008). *A survey of dynamic scheduling in manufacturing systems.* Journal of Scheduling, 12(4), 417–431.

[8] Silver, E. A., Victor, R., Vidal, V., & de Werra, D. (1980). *A tutorial on heuristic methods.* European Journal of Operational Research, 5(3), 153–162.

[9] Juan, A. A., Keenan, P., Martí, R., McGarraghy, S., Panadero, J., Carroll, P., & Oliva, D. (2023). *A review of the role of heuristics in stochastic optimisation: From metaheuristics to learnheuristics.* Annals of Operations Research, 320(2), 831–861.

[10] Lundgren, J., Rönnqvist, M., & Värbrand, P. (2010). *Optimization.* Studentlitteratur AB.

[11] Petropoulos, F., Laporte, G., Aktas, E., Alumur, S. A., Archetti, C., Ayhan, H., Battarra, M., Bennell, J. A., Bourjolly, J.-M., Boylan, J. E., et al. (2023). *Operational research: Methods and applications.* arXiv preprint arXiv:2303.14217.

[12] Youn, S., Geismar, H. N., & Pinedo, M. (2022). Planning and scheduling in healthcare for better care coordination: Current understanding, trending topics, and future opportunities. *Production and Operations Management*, 31(12), 4407–4423.

[13] Liu, H., Wang, Y., & Hao, J. (2024). Solving the patient admission scheduling problem using constraint aggregation. *European Journal of Operational Research*, 316, 85–99.

[14] Demeester, P., Souffriau, W., De Causmaecker, P., & Van den Berghe, G. (2010). A hybrid tabu search algorithm for automatically assigning patients to beds. *Artificial Intelligence in Medicine*, 48(1), 61–70.

[15] Bamigbola, A. F., Bolaji, A. L., Adewole, L. B., Obayomi, A. A., Salaudeen, L. O., & Fagbuagun, O. A. (2023). A hybrid Hill-ABC algorithm for patient admission scheduling problem. *Scientific African.*

[16] Abdalkareem, Z. A., Al-betar, M. A., Amir, A., Ehkan, P., Hammouri, A. I., & Salman, O. H. (2021). Discrete flower pollination algorithm for patient admission scheduling problem. *Computers in Biology and Medicine*, 105007.

[17] Burke, E. K., De Causmaecker, P., Van den Berghe, G., & Van Landeghem, H. (2004). *The State of the Art of Nurse Rostering.* Journal of Scheduling, 7(6), 441–499.

[18] Jung, K. S., Pinedo, M., Sriskandarajah, C., & Tiwari, V. (2019). *Scheduling Elective Surgeries with Emergency Patients at Shared Operating Rooms.* Production and Operations Management, 28, 1407–1430.

[19] Tsang, M. Y., Shehadeh, K., Curtis, F. E., Hochman, B., & Brentjens, T. (2022). *Stochastic Optimization Approaches for an Operating Room and Anesthesiologist Scheduling Problem.* Operational Research.

[20] Ngoo, C. M., Goh, S. L., Sze, S. N., Sabar, N. R., Abdullah, S., & Kendall, G. (2022). *A Survey of the Nurse Rostering Solution Methodologies: The State-of-the-Art and Emerging Trends.* IEEE Access, 10, 56504–56524.

[21] Rajeswari, M., Amudhavel, J., Pothula, S., & Dhavachelvan, P. (2017). *Directed bee colony optimization algorithm to solve the nurse rostering problem.* Computational Intelligence and Neuroscience, 2017, 1–26.

[22] Wickert, T. I., Smet, P., & Van den Berghe, G. (2019). *The nurse rerostering problem: Strategies for reconstructing disrupted schedules.* Computers & Operations Research, 104, 319–337.

[23] Smet, P., De Causmaecker, P., Bilgin, B., & Van den Berghe, G. (2013). *Nurse rostering: A complex example of personnel scheduling with perspectives.* In *Automated Scheduling and Planning.* Berlin, Germany: Springer.

[24] Johansen, A.-S., Nag, B., & Tveit, H. H. (2023). *Nurse Scheduling and Rescheduling: Combining Optimization with Machine Learning-Driven Demand Predictions.* Master's thesis, Norwegian University of Science and Technology, Faculty of Economics and Management, Dept. of Industrial Economics and Technology Management.

[25] Zanda, S., Zuddas, P., & Seatzu, C. (2018). *Long term nurse scheduling via a decision support system based on linear integer programming: A case study at the university hospital in Cagliari.* Computers & Industrial Engineering, 126, 337–347.

[26] Legrain, A., Omer, J., & Rosat, S. (2020). *A rotation-based branch-and-price approach for the nurse scheduling problem.* Mathematical Programming Computation, 12(3), 417–450. doi:10.1007/s12532-019-00172-4. https://hal.science/hal-01545421v2

[27] Legrain, A., & Omer, J. (2024). *A dedicated pricing algorithm to solve a large family of nurse scheduling problems with branch-and-price.* INFORMS Journal on Computing. doi:10.1287/ijoc.2023.0019. https://hal.science/hal-03964952

[28] Boyer, V., Gendron, B., & Rousseau, L.-M. (2013). *A branch-and-price algorithm for the multi-activity multi-task shift scheduling problem.* Journal of Scheduling, 17(2), 185–197. doi:10.1007/s10951-013-0338-9

[29] Rousseau, L.-M., & van Hoeve, W.-J. (2014). *Special issue on: constraint programming approaches to combinatorial optimization.* EURO Journal on Computational Optimization, 2(3), 87–88. doi:10.1007/s13675-014-0027-2

[30] Lim, G. J., Mobasher, A., & Côté, M. J. (2012). *Multi-objective Nurse Scheduling Models with Patient Workload and Nurse Preferences.* Management Science, 2, 149–160.

[31] Ang, B. Y., Lam, S. W. S., Pasupathy, Y., & Ong, M. E. H. (2018). *Nurse workforce scheduling in the emergency department: A sequential decision support system considering multiple objectives*. Journal of Nursing Management, 26(4), 432–441.

[32] Bai, R., Burke, E. K., Kendall, G., Li, J., & McCollum, B. (2010). *A Hybrid Evolutionary Approach to the Nurse Rostering Problem*. IEEE Transactions on Evolutionary Computation, 14, 580–590.

[33] Burke, E. K., & Curtois, T. (2014). *New approaches to nurse rostering benchmark instances*. European Journal of Operational Research, 237(1), 71–81. https://doi.org/10.1016/j.ejor.2014.01.039

[34] Aydas, O. T., Ross, A. D., Scanlon, M. C., & Aydas, B. (2022). *Short-Term Nurse Schedule Adjustments under Dynamic Patient Demand*. Journal of the Operational Research Society, 74, 310–329.

[35] Burke, E. K., Li, J., & Qu, R. (2010). A hybrid model of integer programming and variable neighbourhood search for highly-constrained nurse rostering problems. *European Journal of Operational Research*, 203(2), 484–493.

[36] Awadallah, M. A., Al-Betar, M. A., Khader, A. T., Bolaji, A. L., & Alkoffash, M. (2017). Hybridization of harmony search with hill climbing for highly constrained nurse rostering problem. *Neural Computing and Applications*, 28(3), 463–482.

[37] Jin, S. H., Yun, H. Y., Jeong, S. J., & Kim, K. S. (2017). Hybrid and cooperative strategies using harmony search and artificial immune systems for solving the nurse rostering problem. *Sustainability*, 9(7), 1090.

[38] Chen, P.-S., & Zeng, Z.-Y. (2020). Developing two heuristic algorithms with metaheuristic algorithms to improve solutions of optimization problems with soft and hard constraints: An application to nurse rostering problems. *Applied Soft Computing*, 93, 106336.

[39] Böðvarsdóttir, E. B., Smet, P., Van den Berghe, G., & Stidsen, T. J. R. (2021). Achieving compromise solutions in nurse rostering by using automatically estimated acceptance thresholds. *European Journal of Operational Research*, 292(3), 980–995.

[40] Ingels, J., & Maenhout, B. (2017). Employee substitutability as a tool to improve the robustness in personnel scheduling. *OR Spectrum*, 39(3), 623–658. https://doi.org/10.1007/s00291-017-0476-0

[41] Otero-Caicedo, R., Montoya Casas, C. E., Barajas Jaimes, C., Guzmán Garzón, C. F., Yáñez Vergel, E. A., & Zabala Valdés, J. C. (2023). A preventive–reactive approach for nurse scheduling considering absenteeism and nurses' preferences. *Department of Industrial Engineering, Pontificia Universidad Javeriana*, Bogotá, Colombia.

[42] Maenhout, B., & Vanhoucke, M. (2013). Reconstructing nurse schedules: Computational insights in the problem size parameters. *Omega*, 41(5), 903–918. https://doi.org/10.1016/j.omega.2012.10.010

[43] Wright, P. D., & Mahar, S. M. (2013). Quick-response methods in nurse scheduling to improve hospital performance and reduce disruptions. *European Journal of Operational Research*, 283(1), 390–403. https://doi.org/10.1016/j.ejor.2019.10.047

[44] Burke, E. K., De Causmaecker, P., Berghe, G. V., & Van Landeghem, H. (2004). The state of the art of nurse rostering. *Journal of Scheduling*, 7(6), 441–499.

[45] Wickert, T. I., Smet, P., & Vanden Berghe, G. (2019). The nurse rerostering problem: Strategies for reconstructing disrupted schedules. *Computers & Operations Research*, 104, 319–337. https://doi.org/10.1016/j.cor.2018.12.014

[46] Schoenfelder, J., Bretthauer, K. M., Wright, P. D., & Coe, E. (2020). Nurse scheduling with quick-response methods: Improving hospital performance, nurse workload, and patient experience. *European Journal of Operational Research*, 283(1), 390–403. https://doi.org/10.1016/j.ejor.2019.10.047

[47] Long, Z., Wen, X., Lan, M., & Yang, Y. (2021). Nursing rescheduling problem with multiple rescheduling methods under uncertainty. *Complex & Intelligent Systems*, 1–13. https://doi.org/10.1007/S40747-021-00554-Z

[48] Grolemund, G., & Wickham, H. (2014). R for Data Science: Visualize, Model, Transform, Tidy, and Import Data. *O'Reilly Media.*

[49] Marron, J. S., & Alonso, A. M. (2014). Object oriented data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 6(6), 447–456. https://doi.org/10.1002/wics.1311

[50] Hinterberger, H. (2012). Exploratory data analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1. https://doi.org/10.1002/wics.1234

[51] Lecompte, M. D. (2000). Analyzing qualitative data. *Theory Into Practice*, 39(3), 146–154. https://doi.org/10.1207/s15430421tip3903_5

[52] Capan, M., Khojandi, A., Denton, B. T., Williams, K. D., Ayer, T., Chhatwal, J., Kurt, M., Lobo, J. M., Roberts, M. S., Zaric, G. S., Zhang, S., & Schwartz, J. S. (2017). From data to improved decisions: Operations research in healthcare delivery. *Medical Decision Making*, 37, 849–859. https://doi.org/10.1177/0272989X17729327

[53] Carter, M. W., Hans, E. W., & Kolisch, R. (2012). Health care operations management. *Health Care Management Science*, 15, 1–3. https://doi.org/10.1007/s10729-012-9208-5

[54] Denton, B. T. (2023). Frontiers of medical decision-making in the modern age of data analytics. *IISE Transactions*, 55, 94–105. https://doi.org/10.1080/24725854.2022.2124189

[55] Baardman, L., Cristian, R., Perakis, G., Singhvi, D., SkaliLami, O., & Thayaparan, L. (2022). The role of optimization in some recent advances in data-driven decision-making. *IISE Transactions*, Published online August 11, 2022. https://doi.org/10.1080/24725854.2022.2109471

[56] Behrens, J. T. (1997). *Principles and procedures of exploratory data analysis.* Psychological Methods, 2(2), 131–160. doi:10.1037/1082-989x.2.2.131

[57] Chatfield, C. (1986). *Exploratory data analysis.* European Journal of Operational Research, 23(1), 5–13. doi:10.1016/0377-2217(86)90209-2

[58] Watson, J. C. (2017). *Establishing Evidence for Internal Structure Using Exploratory Factor Analysis.* Measurement and Evaluation in Counseling and Development, 50(4), 232–238. doi:10.1080/07481756.2017.1336931

[59] Smith-Miles, K. A. (2012). *Cross-disciplinary perspectives on meta-learning for algorithm selection.* Computing Surveys, 41(1), Article 6.

[60] Michard, F., & Teboul, J.-L. (2019). *Predictive analytics: beyond the buzz.* Annals of Intensive Care, 9(1), 1–3. doi:10.1186/s13613-019-0524-9

[61] Rodrigues, L., & Givigi, S. N. (2024). *Predictive Analytics: an Optimization Perspective.* IEEE Access, 12, 1–15.

[62] Lu, Y., Garcia, R., Hansen, B., Gleicher, M., & Maciejewski, R. (2017). *The State-of-the-Art in Predictive Visual Analytics.* Computer Graphics Forum, 36(3), 539–562. doi:10.1111/cgf.13210

[63] Hofstetter, H., Dusseldorp, E., Zeileis, A., & Schuller, A. A. (2016). *Modeling Caries Experience: Advantages of the Use of the Hurdle Model.* Caries Research, 50(6), 517–526. doi:10.1159/000448197

[64] Bethell, J., Rhodes, A. E., Bondy, S. J., Lou, W. Y. W., & Guttmann, A. (2010). *Repeat self-harm: application of hurdle models.* British Journal of Psychiatry, 196(03), 243–244. doi:10.1192/bjp.bp.109.068809

[65] Brilleman, S., Crowther, M., May, M. T., Gompels, M., & Abrams, K. (2016). *Joint longitudinal hurdle and time-to-event models: an application related to viral load and duration of the first treatment regimen in HIV patients initiating therapy.* Statistics in Medicine, 35(20), 3583–3594. https://doi.org/10.1002/sim.6948

[66] Moffatt, P. G. (2005). *Hurdle models of loan default.* Journal of the Operational Research Society, 56(9), 1063–1071. doi:10.1057/palgrave.jors.2601922

[67] Feng, C. X. (2021). *A comparison of zero-inflated and hurdle models for modeling zero-inflated count data.* Journal of Statistical Distributions and Applications, 8(1), 8. doi:10.1186/s40488-021-00123-2

[68] Mutiso, F., Pearce, J. L., Benjamin-Neelon, S. E., Mueller, N. T., Li, H., & Neelon, B. (2021). *Bayesian negative binomial regression with spatially varying dispersion: Modeling COVID-19 incidence in Georgia.* Spatial and Spatio-temporal Epidemiology, 39, 100442.

[69] Hu, B., Shao, J., & Palta, M. (2006). *Pseudo-R^2 in logistic regression model.* Statistica Sinica, 16, 847–860.

[70] Google Developers. (2025). *Classification: ROC et AUC – Machine Learning Crash Course.* Consulted from https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc

Appendices

Appendix A

Comprehensive Absence Analysis

1.1 Comprehensive Absence Analysis for Section 1

Total Days Analyzed by Weekday

Weekday	**Total Days**
Monday	261
Tuesday	261
Wednesday	261
Thursday	261
Friday	261
Saturday	261
Sunday	260

Table A.1: Total number of days analyzed for all shifts by weekday

1.1.1 Morning Shift

Weekday	0	1	2	3	4	5	6	7	8
Monday	104	75	52	20	6	4	0	0	0
Tuesday	78	75	66	26	12	4	0	0	0
Wednesday	78	72	56	36	11	5	1	1	1
Thursday	85	71	55	31	13	5	1	0	0
Friday	78	79	55	27	15	4	3	0	0
Saturday	150	72	27	10	2	0	0	0	0
Sunday	132	88	29	8	3	0	0	0	0

Table A.2: Count of days by number of morning absences

Weekday	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0
Monday	0.398	0.287	0.199	0.077	0.023	0.015	0.000	0.000	0.000
Tuesday	0.299	0.287	0.253	0.100	0.046	0.015	0.000	0.000	0.000
Wednesday	0.299	0.276	0.215	0.138	0.042	0.019	0.004	0.004	0.004
Thursday	0.326	0.272	0.211	0.119	0.050	0.019	0.004	0.000	0.000
Friday	0.299	0.303	0.211	0.103	0.057	0.015	0.011	0.000	0.000
Saturday	0.575	0.276	0.103	0.038	0.008	0.000	0.000	0.000	0.000
Sunday	0.508	0.338	0.112	0.031	0.012	0.000	0.000	0.000	0.000

Table A.3: Frequency of morning absences

Weekday	**Criticality Score**
Monday	1.084
Tuesday	1.352
Wednesday	1.464
Thursday	1.368
Friday	1.410
Saturday	0.628
Sunday	0.700

Table A.4: Criticality scores by weekday for morning shift

1.1.2 Evening Shift

Weekday	0	1	2	3	4	5
Monday	95	99	53	12	0	2
Tuesday	112	93	40	13	2	1
Wednesday	94	97	47	18	5	0
Thursday	111	96	36	15	3	0
Friday	152	73	27	7	2	0
Saturday	141	79	30	9	2	0
Sunday	139	83	25	12	1	0

Table A.5: Count of days by number of evening absences

Weekday	0.0	1.0	2.0	3.0	4.0	5.0
Monday	0.364	0.379	0.203	0.046	0.000	0.008
Tuesday	0.429	0.356	0.153	0.050	0.008	0.004
Wednesday	0.360	0.372	0.180	0.069	0.019	0.000
Thursday	0.425	0.368	0.138	0.057	0.011	0.000
Friday	0.582	0.280	0.103	0.027	0.008	0.000
Saturday	0.540	0.303	0.115	0.034	0.008	0.000
Sunday	0.535	0.319	0.096	0.046	0.004	0.000

Table A.6: Frequency of evening absences

Weekday	**Criticality Score**
Monday	0.962
Tuesday	0.862
Wednesday	1.015
Thursday	0.862
Friday	0.598
Saturday	0.667
Sunday	0.665

Table A.7: Criticality scores by weekday for evening shift

1.1.3 Night Shift

Weekday	0	1	2	3	4
Monday	181	60	17	3	0
Tuesday	161	71	20	9	0
Wednesday	143	89	25	3	1
Thursday	154	75	28	4	0
Friday	142	92	23	3	1
Saturday	175	57	27	2	0
Sunday	173	56	28	3	0

Table A.8: Count of days by number of night absences

Weekday	0.0	1.0	2.0	3.0	4.0
Monday	0.693	0.230	0.065	0.011	0.000
Tuesday	0.617	0.272	0.077	0.034	0.000
Wednesday	0.548	0.341	0.096	0.011	0.004
Thursday	0.590	0.287	0.107	0.015	0.000
Friday	0.544	0.352	0.088	0.011	0.004
Saturday	0.670	0.218	0.103	0.008	0.000
Sunday	0.665	0.215	0.108	0.012	0.000

Table A.9: Frequency of night absences

Weekday	**Criticality Score**
Monday	0.395
Tuesday	0.529
Wednesday	0.582
Thursday	0.548
Friday	0.579
Saturday	0.448
Sunday	0.465

Table A.10: Criticality scores by weekday for night shift

1.2 Comprehensive Absence Analysis for Section 2

1.2.1 Morning Shift

Weekday	0	1	2	3	4	5	6	7
Monday	102	86	41	26	5	1	0	0
Tuesday	98	75	52	20	12	2	2	0
Wednesday	79	82	51	26	18	4	1	0
Thursday	101	72	40	34	8	5	1	0
Friday	84	80	57	25	7	6	0	2
Saturday	156	78	24	3	0	0	0	0
Sunday	135	91	25	9	0	0	0	0

Table A.11: Count of days by number of morning absences (Section 2)

Weekday	0.0	1.0	2.0	3.0	4.0	5.0	6.0	7.0
Monday	0.391	0.330	0.157	0.100	0.019	0.004	0.000	0.000
Tuesday	0.375	0.287	0.199	0.077	0.046	0.008	0.008	0.000
Wednesday	0.303	0.314	0.195	0.100	0.069	0.015	0.004	0.000
Thursday	0.387	0.276	0.153	0.130	0.031	0.019	0.004	0.000
Friday	0.322	0.307	0.218	0.096	0.027	0.023	0.000	0.008
Saturday	0.598	0.299	0.092	0.011	0.000	0.000	0.000	0.000
Sunday	0.519	0.350	0.096	0.035	0.000	0.000	0.000	0.000

Table A.12: Frequency of morning absences (Section 2)

Weekday	**Criticality Score**
Monday	1.038
Tuesday	1.184
Wednesday	1.379
Thursday	1.215
Friday	1.307
Saturday	0.517
Sunday	0.646

Table A.13: Criticality scores by weekday for morning shift (Section 2)

1.2.2 Evening Shift

Weekday	0	1	2	3	4
Monday	128	91	29	13	0
Tuesday	116	76	50	17	2
Wednesday	104	96	41	15	5
Thursday	118	98	35	10	0
Friday	153	77	26	4	1
Saturday	128	92	32	9	0
Sunday	151	80	25	4	0

Table A.14: Count of days by number of evening absences (Section 2)

Weekday	0.0	1.0	2.0	3.0	4.0
Monday	0.490	0.349	0.111	0.050	0.000
Tuesday	0.444	0.291	0.192	0.065	0.008
Wednesday	0.398	0.368	0.157	0.057	0.019
Thursday	0.452	0.375	0.134	0.038	0.000
Friday	0.586	0.295	0.100	0.015	0.004
Saturday	0.490	0.352	0.123	0.034	0.000
Sunday	0.581	0.308	0.096	0.015	0.000

Table A.15: Frequency of evening absences (Section 2)

Weekday	**Criticality Score**
Monday	0.720
Tuesday	0.900
Wednesday	0.931
Thursday	0.759
Friday	0.556
Saturday	0.701
Sunday	0.546

Table A.16: Criticality scores by weekday for evening shift (Section 2)

1.2.3 Night Shift

Weekday	0	1	2	3
Monday	194	58	7	2
Tuesday	147	93	20	1
Wednesday	148	90	22	1
Thursday	159	83	18	1
Friday	165	78	15	3
Saturday	179	70	9	3
Sunday	176	69	12	3

Table A.17: Count of days by number of night absences (Section 2)

Weekday	0.0	1.0	2.0	3.0
Monday	0.743	0.222	0.027	0.008
Tuesday	0.563	0.356	0.077	0.004
Wednesday	0.567	0.345	0.084	0.004
Thursday	0.609	0.318	0.069	0.004
Friday	0.632	0.299	0.057	0.011
Saturday	0.686	0.268	0.034	0.011
Sunday	0.677	0.265	0.046	0.012

Table A.18: Frequency of night absences (Section 2)

Weekday	**Criticality Score**
Monday	0.299
Tuesday	0.521
Wednesday	0.525
Thursday	0.467
Friday	0.448
Saturday	0.372
Sunday	0.392

Table A.19: Criticality scores by weekday for night shift (Section 2)

1.3 Comprehensive Absence Analysis for Section 3

1.3.1 Morning Shift

Weekday	0	1	2	3	4	5	6	9
Monday	75	70	49	40	23	3	1	0
Tuesday	63	69	66	39	14	9	1	0
Wednesday	52	57	75	41	24	9	3	0
Thursday	72	60	74	34	13	6	1	1
Friday	41	52	62	66	24	6	10	0
Saturday	179	62	17	3	0	0	0	0
Sunday	154	79	22	5	0	0	0	0

Table A.20: Count of days by number of morning absences (Section 3)

Weekday	0.0	1.0	2.0	3.0	4.0	5.0	6.0	9.0
Monday	0.287	0.268	0.188	0.153	0.088	0.011	0.004	0.000
Tuesday	0.241	0.264	0.253	0.149	0.054	0.034	0.004	0.000
Wednesday	0.199	0.218	0.287	0.157	0.092	0.034	0.011	0.000
Thursday	0.276	0.230	0.284	0.130	0.050	0.023	0.004	0.004
Friday	0.157	0.199	0.238	0.253	0.092	0.023	0.038	0.000
Saturday	0.686	0.238	0.065	0.011	0.000	0.000	0.000	0.000
Sunday	0.592	0.304	0.085	0.019	0.000	0.000	0.000	0.000

Table A.21: Frequency of morning absences (Section 3)

Weekday	**Criticality Score**
Monday	1.536
Tuesday	1.628
Wednesday	1.874
Thursday	1.559
Friday	2.146
Saturday	0.402
Sunday	0.531

Table A.22: Criticality scores by weekday for morning shift (Section 3)

1.3.2 Evening Shift

Weekday	0	1	2	3	4
Monday	113	99	34	14	1
Tuesday	135	81	41	3	1
Wednesday	116	98	38	9	0
Thursday	97	91	57	16	0
Friday	161	72	24	4	0
Saturday	144	84	29	4	0
Sunday	161	68	27	3	1

Table A.23: Count of days by number of evening absences (Section 3)

Weekday	0.0	1.0	2.0	3.0	4.0
Monday	0.433	0.379	0.130	0.054	0.004
Tuesday	0.517	0.310	0.157	0.011	0.004
Wednesday	0.444	0.375	0.146	0.034	0.000
Thursday	0.372	0.349	0.218	0.061	0.000
Friday	0.617	0.276	0.092	0.015	0.000
Saturday	0.552	0.322	0.111	0.015	0.000
Sunday	0.619	0.262	0.104	0.012	0.004

Table A.24: Frequency of evening absences (Section 3)

Weekday	**Criticality Score**
Monday	0.816
Tuesday	0.674
Wednesday	0.770
Thursday	0.969
Friday	0.506
Saturday	0.590
Sunday	0.519

Table A.25: Criticality scores by weekday for evening shift (Section 3)

1.3.3 Night Shift

Weekday	0	1	2	3
Monday	225	34	2	0
Tuesday	223	36	2	0
Wednesday	212	44	4	1
Thursday	217	37	6	1
Friday	221	35	4	1
Saturday	216	43	2	0
Sunday	221	37	2	0

Table A.26: Count of days by number of night absences (Section 3)

Weekday	0.0	1.0	2.0	3.0
Monday	0.862	0.130	0.008	0.000
Tuesday	0.854	0.138	0.008	0.000
Wednesday	0.812	0.169	0.015	0.004
Thursday	0.831	0.142	0.023	0.004
Friday	0.847	0.134	0.015	0.004
Saturday	0.828	0.165	0.008	0.000
Sunday	0.850	0.142	0.008	0.000

Table A.27: Frequency of night absences (Section 3)

Weekday	**Criticality Score**
Monday	0.146
Tuesday	0.153
Wednesday	0.211
Thursday	0.199
Friday	0.176
Saturday	0.180
Sunday	0.158

Table A.28: Criticality scores by weekday for night shift (Section 3)

Weekly Criticality Comparison by Shift

Weekday	**Morning**	**Evening**	**Night**
Monday	1.536	0.816	0.146
Tuesday	1.628	0.674	0.153
Wednesday	1.874	0.770	0.211
Thursday	1.559	0.969	0.199
Friday	2.146	0.506	0.176
Saturday	0.402	0.590	0.180
Sunday	0.531	0.519	0.158

Table A.29: Criticality scores comparison by shift (Section 3)