ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

A SIMULATION-BASED STAFF SCHEDULING ANALYSIS FOR CALL CENTERS IN CARGO INDUSTRY

M.Sc. THESIS

Res. Asst. Ömer Faruk KANDAZ

Faculty of Management

Management Engineering Programme

JUNE 2024

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

A SIMULATION-BASED STAFF SCHEDULING ANALYSIS FOR CALL CENTERS IN CARGO INDUSTRY

M.Sc. THESIS

Res. Asst. Ömer Faruk KANDAZ
(507201060)

Faculty of Management

Management Engineering Programme

Thesis Advisor: Assoc. Prof. Tuncay ÖZCAN

JUNE 2024

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

KARGO SEKTÖRÜNDE ÇAĞRI MERKEZLERİ İÇİN SİMÜLASYON TABANLI PERSONEL ÇİZELGELEME ANALİZİ

YÜKSEK LİSANS TEZİ

Res. Asst. Ömer Faruk KANDAZ
(507201060)

İşletme Mühendisliği Anabilim Dalı

İşletme Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Assoc. Prof. Tuncay ÖZCAN

HAZİRAN 2024

Res. Asst. Ömer Faruk KANDAZ, a M.Sc. student of ITU Graduate School student ID 507201060 successfully defended the thesis entitled “A SIMULATION-BASED STAFF SCHEDULING ANALYSIS FOR CALL CENTERS IN CARGO INDUSTRY”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Assoc. Prof. Tuncay ÖZCAN** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Hatice Camgöz AKDAĞ** Istanbul Technical University	....
	Assoc. Prof. Ersin NAMLI Istanbul University	....
		....

Date of Submission : **20 May 2024**
Date of Defense : **07 June 2024**

To my spouse and children,

FOREWORD

For foreword, in the depths of gratitude, I extend my profound thanks to my family, who unwaveringly believed in me without hesitation. They imparted upon me the priceless value of learning, not merely once, but throughout my journey. By placing their trust in me, they fashioned the very essence of my being. My greatest strength lies in finding joy amidst life's trials, a spirit that propels me to conquer all. Hence, among my indebtedness, my family reigns supreme. No words can adequately convey the depth of my gratitude to them.

The path to success is not innate; it is forged through the unwavering support of those who stand by you. Thus, I am deeply grateful to my thesis advisor, Assoc. Prof. Tuncay Özcan, for his unyielding encouragement and invaluable guidance. He exemplifies the epitome of mentorship, sharing his profound wisdom while fostering an environment conducive to learning. His tutelage has empowered me to envision, innovate, and strive for excellence in the realm of academia.

Furthermore, the academic realm beckons to those with a fervent passion akin to a blazing flame, and to those who tirelessly pursue their aspirations. I have always harbored such a flame within me, yet without a spark, it remains dormant. Hence, I extend heartfelt appreciation to Prof. Dr. Hatice Camgöz Akdağ for igniting that spark and guiding my path.

The pursuit of knowledge is boundless, akin to a distant horizon that may seem unattainable to a single soul. Yet, it is the journey itself, marked by steadfast determination, that heralds true success. Therefore, I conclude by reiterating my profound gratitude to both of my esteemed professors, who have indelibly shaped my character and fueled my aspirations.

June 2024 Res. Asst. Ömer Faruk KANDAZ
(Research Assistant)

TABLE OF CONTENTS

ABBREVIATIONS

ABC	: Artificial Bee Colony
ACD/CTI	: Automatic Call Distribution / Computer Telephony Integration
APS	: Augmented Probability Simulation
ARIMA	: Autoregressive Integrated Moving Average
ASA	: Average Answering Speed
B2B	: Business-to-Business
BBC	: Business-to-Consumer
BIP	: Binary Integer Programming
CTMC	: Continuous Time Markov Chain
DGA	: Distributed Genetic Algorithm
DES	: Discrete-Event Simulation
EABC	: Enhanced Artificial Bee Colony
EEJS	: Energy Efficient Job Scheduling
GA	: Genetic Algorithm
GISG	: Generate, Improve, Select, Generate
HABC	: Hybrid Artificial Bee Colony
IDEF	: Integration Definition Modeling
IG	: Iterated Greedy
CTMC	: Continuous Time Markov Chain
IP	: Integer Programming
ISA	: Iterative Staffing Algorithm
KPI	: Key Performance Indicator
LP	: Linear Programming
MILP	: Mixed-Integer Linear Programming
MICN	: Mean Incoming Call Numbers
MIP	: Mıxed-Integer Programming
NHPP	: Nonhomogeneous Poisson Process
NP	: Non-Polynomial
CTMC	: Continuous Time Markov Chain
SA	: Simulated Annealing
SAA	: Sample Average Approximation
SACCPM	: Simulation-Based Analytic Center Cutting Plane Method
SAH	: Simulated Annealing Heuristic
SIP	: Stochastic Integer Programming
SLA	: Service Level Agreement
SMPTSP	: Shift Minimization Personnel Task Scheduling Problem

SYMBOLS

A	: Arrival process
D	: Queue's discipline
K	: Number of places
M	: Markovian arrival process
N	: Average number of jobs in the system
S	: Service time distribution
W	: Average time a jobs spends in the system
$\mathbf{Q_t}$	: State Vector
$\mathbf{S_t}$	: System State
$\mathbf{X_t}$	: Input Trajectory Vector
c	: Number of servers
n	: Number of trials
t	: Time
x	: Number of successes
λ	: Number of events

LIST OF TABLES

LIST OF FIGURES

A SIMULATION-BASED STAFF SCHEDULING ANALYSIS FOR CALL CENTERS IN CARGO INDUSTRY

SUMMARY

Call center staff scheduling problems have been studied frequently on a regular basis for many years. Call centers generally conduct their business on customers by delivering service. Problems occur especially when the density of queues is high, causing long waiting times. In terms of customer satisfaction, this outcome is undesirable.

One of the reasons that causes this outcome is the inefficient personnel/job scheduling. In call centers, most of the time, the demand of service is inhomogeneous. Therefore, the incoming number of calls varies depending on the time interval of the day, in which the firm operates. The aim of this thesis is to understand appropriately how a call center operates and generate effective and efficient solutions by employing simulation.

The raw data regarding the call center operations were first acquired which contained several important aspects that are beneficial to capture the behavior of the system and propose a better alternative to it. Gathered raw data had incoming call dates and several other statistical records that were kept.

Firstly, this data was analyzed using MATLAB. Subsequent calculations on Matlab made it possible to put the resulting data that illustrate different aspects of the queue to the ARENA simulation program as inputs. To achieve effective and efficient scheduling, the incoming calls were categorized based on different time intervals by taking the density of calls into account across a day for 27 days.

Afterward, input analysis was conducted on the ARENA Input Analyzer tool to get the best-fit distributions that were necessary in order to create a successful simulation analysis.

After processes were clearly defined and mapped on ARENA, several simulation parameters were calculated and considered before conducting the analysis, one of which stands for huge importance in terms of reliability and accuracy of the outcomes, which was determining the appropriate number of replications. The results showed that there was insufficient number of employees working at that time. Additionally, employee scheduling was not conducted in an efficient way. These were the main reasons why bottlenecks occurred in the first place. In order to reinforce the solutions, a scenario analysis was conducted. Additional three scenarios were included in the simulation program and all of the scenarios were compared against each other in terms of KPIs.

Finally, results showed that the simulation conducted appeared to be appropriate, and represented the overall system successfully. The scenario analysis illustrated that

relaxing the bottleneck could be achieved when effective and efficient job scheduling is conducted. Moreover, the outcomes also indicated that customer satisfaction could be achieved by improving the existing system according to the simulation results presented.

In the first section, an entry to the call center staff scheduling problem has been made. In section 2, a literature study was conducted in order to see the span and spread of techniques employed to solve the problem. A comprehensive study was conducted to see the blindspots in literature. These studies were heavily analyzed, and the outcomes were clearly illustrated. It was deducted from these studies that the use of simulation would be appropriate, specifically, discrete-event simulation.

In Section 3, the methodology is proposed. After determining which technique to employ in order to solve the problem, related academic books were analyzed to come up with a fine approach to the problem.

In Section 4, after creating the framework to follow, the simulation application was conducted. Firstly, the raw data was pre-processed and analyzed in Matlab, the descriptive statistics were presented, and the results were given to the ARENA simulation program to conduct input analysis.

Input analysis included conducting data analysis, and determining best-fit distributions. Data analysis and finding best-fit distributions are crucially important because the ultimate aim when conducting a simulation analysis is to create an artificial system that perfectly reflects the original one. Because the raw data acquired showed seasonality and, therefore needed to be more carefully analyzed. Best-fit distributions were found for the five main data categories which were call durations, interarrival times, waiting time for missed calls, assignment time to the queue, and connection waiting time from the queue to the operator.

The application chapter also included creating a process flow chart. After the system is conceptualized and input analysis is conducted, the process flow chart, which also reflects the actual simulation model that needs to be built, was drawn. Subsequently, the ARENA model was created.

In the ARENA model, along with several other aspects, one main parameter was to define the minimum necessary number of replications. In order to do that, some statistical calculations and comparisons were made, and the minimum necessary number of replications was found.

Finally, output and scenario analysis were conducted to better capture the the effects of changes made in order to improve the overall system. Important KPIs were collected and run for four scenarios. It seemed that waiting time in queues was caused by the insufficient number of employees working without an appropriate schedule.

Additionally, this caused the problem of long waiting times in queues, in other words, bottlenecks, which also decreased the customer satisfaction level. Scenario analysis pointed out the existing system and proposed better alternatives in the subsequent scenarios in terms of KPIs.

According to the results of the scenario analysis, the firm needs to hire more personnel and schedule those effectively and efficiently, as illustrated in this thesis to improve the overall call center operations.

All the necessary figures and tables regarding the simulation application were given in Section 4.

Additionally, in Section 5, the results of this thesis were presented and suggestions for future studies were given.

**KARGO SEKTÖRÜNDE ÇAĞRI MERKEZLERİ
İÇİN SİMÜLASYON TABANLI
PERSONEL ÇİZELGELEME ANALİZİ**

ÖZET

Çağrı merkezi personelinin çizelgeleme problemleri düzenli olarak uzun yıllardır sıklıkla araştırılmaktadır. Çağrı merkezleri, genellikle işlerini müşterilerine hizmet sağlayarak yürütmektedir.

Çağrı merkezlerinde çoğunlukla yaşanan en önemli problemi uzun bekleme süreleri, yani telefon hattında oluşan kuyruklar oluşturmaktadır.

Kuyruk yoğunluğunun artması, beraberinde uzun bekleme süreleri gibi problemleri de getirmektedir. Bu durum, müşteri memnuniyetini sağlamaya çalışan bir firma için istenmeyen durumlar kategorisine girer. Çünkü uzun bekleme süreleri, daha fazla beklemeden çağrıyı sonlandırma kararına neden olabilmektedir.

Günün en yoğun saatlerinde, kuyruk hattının uzaması doğal olduğu gibi, personelin iş çizelgelemesinin ne kadar doğru yapıldığı da diğer yandan önem arz etmektedir.

Personelin iş çizelgelemesi doğru yapılmadığında, önlenebilecek ya da azaltılabilecek kuyruk yoğunluğu, aksine artarmaktadır.

Çağrı merkezlerinde çoğu zaman hizmet talebi homojen değildir. Bu nedenle, firmanın faaliyet gösterdiği günlerde, gelen çağrı sayısı çağrının zaman aralığına göre değişkenlik göstermektedir.

Bu tezin amacı, bir çağrı merkezinin tam olarak nasıl çalıştığını anlamak ve bu doğrultuda etkin ve verimli yaklaşımları simülasyon kullanarak oluşturmaktır.

Çağrı merkezi operasyonlarına ilişkin ilk olarak ham veriler elde edilir. Bu veriler, çağrı merkezindeki sistem davranışını yakalamada (daha iyi bir alternatif sunabilmek için) faydalı olan birkaç önemli hususu içermektedir.

Toplanan ham veriler, gelen çağrı tarihleri ve tutulan diğer bazı istatistiksel kayıtları da içermiştir. Bu veriler ilk önce Matlab'da analiz edilmiştir.

Ardından Matlab'da yapılan hesaplamalar, kuyruğun farklı yönlerini gösteren verileri ARENA simülasyon programına girdi olarak koyulmasını sağlamıştır.

Etkili ve verimli bir planlama sağlamak için gelen çağrılar, her gün gelen çağrı yoğunluğu 27 gün boyunca dikkate alınarak farklı zaman aralıklarına göre kategorize edilmiştir.

Daha sonra girdi analizi yapılmıştır. ARENA simülasyon programında girdilerin analiz edilebilmesini sağlayan "Input Analyzer" aracı, ihtiyaç duyulan en uygun dağılımları elde etmek ve bunun sonucunda başarılı bir simülasyon analizi oluşturmak için kullanılmıştır.

Süreçler açıkça tanımlandıktan sonra ARENA'da kodlanmış ve sonuçları güvenilirliği ve doğruluğu açısından büyük öneme sahip olan gerekli yineleme sayısı hesaplanmıştır.

Sonuçlar, çalışan personel sayısının yetersiz sayıda olduğunu göstermiştir. Ayrıca personelin iş çizelgelemesi verimli ve etkin bir şekilde yapılmadığı için sistemde darboğaz oluştuğu gözlemlenmiştir.

Bu gözlemlere ek olarak, çözümleri/alternatifleri güçlendirmek amacıyla senaryo analizi yapılmıştır. Simülasyon programına ilave üç senaryo dahil edilmiş ve bu senaryoların tümü KPI'lar açısından birbirleriyle karşılaştırılmıştır.

Elde edilen senaryo analizleri sonucunda, yapılan simülasyonun uygun olduğu ve gerçek sistemi başarıyla ve bire bir yansıttığı tespit edilmiştir.

Buna ek olarak, senaryo analizi, darboğazın hafifletilmesinin, etkin ve verimli iş planlaması yapıldığı takdirde mümkün olabileceğini göstermiştir. Dahası, sonuçlar müşteri memnuniyetinin, bu tezde sunulan simülasyon sonuçları ışığında sistemin iyileştirilmesiyle elde edilebileceğini ortaya koymuştur.

Birinci bölümde çağrı merkezi personel çizelgeleme problemine giriş yapılmıştır. İkinci bölümde ise, kullanılan yöntemlerin kapsam ve yayılımını görmek amacıyla literatür çalışması yapılmıştır. Bu çalışmalar detaylı bir şekilde analiz edildi ve sonuçlar açık bir şekilde gösterilmiştir.

Literatür çalışmalarına bakılarak ayrık olay simülasyonunun kullanılmasının uygun olduğu çıkarımı yapılmıştır.

Üçüncü bölümde, kullanılması planlanan tekniğin, diğer alternatif tekniklerle beraber, detaylı araştırması yapılmıştır ve bu bağlamda çeşitli akademik kitaplar, probleme iyi bir çözüm yaklaşımı geliştirmek amacıyla, detaylı bir şekilde incelenmiştir.

Dördüncü bölümde, takip edilecek yol haritasının çerçevesi belirlendikten sonra simülasyon uygulaması yapılmış. Öncelikle ham verilerin ön analizi Matlab program kullanarak yapıldı ve ham veriye ait tanımlayıcı istatistiksel veriler sunulmuş, ardından bu veriler ARENA simülasyon programına girdi analizinin yapılması amacıyla verilmiştir.

Girdi analizi, veri analizinin yapılmasını ve en uygun dağılımların belirlenmesini içermektedir. Veri analizi ve en uygun dağılımların bulunması hayati önem taşımaktadır, çünkü simülasyon analizi yaparken nihai amaç, orijinalini mükemmel bir şekilde yansıtan, yapay bir sistem oluşturmaktır.

Elde edilen ham veriler mevsimsellik gösterdiği için daha dikkatli analiz edilmesi gerekmektedir.

En uygun dağılımlar, çağrı merkezinden alınan beş ana veri kategorisi göz önünde bulundurularak hesaplanmıştır. Bu kategoriler, çağrı süreleri, gelişler arası süreler, kaçan çağrıların bekleme süresi, kuyruğa atanma süresi ve kuyruktan operatöre bağlanma bekleme süreleridir.

Uygulama bölümünde ayrıca süreç akış şemasının oluşturulması da yer almıştır. Sistem kavramsallaştırılıp girdi analizi yapıldıktan sonra süreç akış şeması oluşturulmuştur. Hemen ardından, ARENA modeli kurulmuştur.

ARENA modelinde diğer bazı hususların yanı sıra bir ana parametre de, gerekli minimum yineleme/tekrarlama sayısını hesaplamaktır. Bunu yapabilmek için bazı istatistiksel hesaplamalar ve karşılaştırmalar yapılmış ve gerekli minimum tekrarlama sayısı tespit edilmiştir.

Son olarak, sistemi genel anlamda iyileştirebilmek adına yapılan değişiklikleri etkilerini yakalayabilmek için senaryo analizleri yapılmıştır. Dört senaryo için önemli anahtar performans göstergeleri hesaplanmıştır. Buna göre, kuyrukta bekleme süresine yetersiz sayıda personelin ve etkin bir şekilde yapılmayan iş çizelgelemesinin neden olduğu sonucuna varılmıştır.

Ek olarak, bu durum, kuyrukta uzun bekleme sürelerinin oluşmasına, diğer bir deyişle darboğaz oluşumuna neden olmuştur. Söz konusu firmanın hedeflerinden biri de müşteri memnuniyetinin sağlanmasıdır. Bu bağlamda, senaryo analizi, var olan sistemi ve sunulan iyileştirme alternatiflerini ortaya koymuştur.

Bu sonuçlara göre, firma daha fazla personeli işe almalı ve bunları bu tezde sunulduğu gibi etkin ve verimli bir şekilde çizelgelemelidir.

Dördüncü bölümde, simülasyon uygulaması ile ilgili bütün gerekli tablo ve şekiller verilmiştir. Buna ek olarak, bazı yönetsel çıktıların altı çizilmiştir.

Bölüm 5'de tezin sonuçları sunulmuş ve gelecek çalışmalar için önerilerde bulunulmuştur.

1. INTRODUCTION

The staff scheduling problem is a common problem often dealt with across different service sectors not only for B2B but also for B2C businesses. Therefore, the problem needs to be clearly understood while demanding a high-quality solution.

Almost most of the business sector that operates specifically in the service sector faces the problem of not being able to schedule their staff effectively and efficiently. Additionally, the issue regarding the problem of not being able to manage their service operations also arises when scheduling is not conducted accurately which causes customer dissatisfaction.

To that end, these two objectives, which are improving the scheduling and meeting customer demands, increase in value.

On the other hand, what stands out as a crucial factor is the technique used to conduct the staff scheduling analysis. There are a number of techniques that can be employed for this study, which fall into the main categories such as optimization, simulation and forecasting.

Another way of looking at it is to employ the hybrid versions of these.

Hybrid versions of these techniques have the potential to capture several aspects of the problem. Moreover, a real-life scheduling problem can be categorized as a complex problem, therefore, employing both simulation and optimization iteratively would yield more reliable results. These studies exist in the literature whereas the optimization techniques employed mostly bounded with deterministic approaches such as linear or integer programming which are also not enough to align with the stochastic nature of the problem.

In this study, a real-life call center staff scheduling problem is dealt with. Incoming call data of a company that -operates in cargo sector- for March 2022 were collected from the firm in order to improve the overall system considering the managerial objectives.

The purpose of this thesis is to generate and propose an effective and efficient solution to the staff scheduling problem in call centers. The literature is analyzed to find out which studies have been conducted and how those resulted.

The ultimate aim consists of several aspects which are mainly increasing customer satisfaction by improving bottlenecks, scheduling staff effectively and efficiently considering managerial aspects, and proposing a better alternative to the current literature by proposing an even better framework that will yield reliable and accurate results when employed.

In short, the purpose of this thesis is to propose a solution that will be considered solid and reliable and offer the framework and solution technique used for future studies in this context.

This thesis comprises a literature review in Section 2, the development of the methodology for the related study in Section 3, the execution of simulation applications in Section 4, and culminates with concluding remarks presented in Section 5.

2. LITERATURE REVIEW

There are various studies in literature on staff scheduling, mostly in different service sectors. Analyzing those is crucial to point out the gaps in the literature. Therefore, in this section, the studies that are mostly related to staff scheduling have been collected, provided, and illustrated.

At the end of this section, there exists a taxonomy table in which the techniques used in literature are extracted and shown.

2.1 Literature Studies on Staff Scheduling Using Optimization

In this section, literature studies conducted on staff scheduling using optimization technique is presented.

Thompson, G. M. [1] studied on a simulated annealing heuristic model to solve labor/staff scheduling problems. The model was coded in FORTRAN and the performance was evaluated using integer programming (IP) model coded in GAMS program. In this study, it was assumed that the staff had homogeneous skills. This model was also tested against 144 test problems. Simulated annealing heuristic (SAH) model were found to be more costly, approximately 0.29 percent, but it was also found efficient in terms of time requirements to come up with an optimal solution. The author suggested that the future work could contain using complementary criteria in order to create neighbourhood in SA.

Easton, F. F., & Mansour, N. [2] analyzed staff scheduling in call centers. The authors developed an algorithm that can be used for three types of problems which are General Set Covering, Deterministic Goal and Stochastic Goal. Distributed genetic algorithm (DGA) was used as optimization method. The developed algorithm was compared against both Simulated annealing (SA) and Tabu search (TS) algorithms. As a result, DGA outperformed both of these metaheuristic algorithms. Moreover, in terms of performance evaluation, IBM/RS 6000 workstations were used. The authors suggested that a possible future work could be on computational and space efficiency of DGA.

Cai, X., & Li, K. N. [3] conducted research on staff scheduling that primarily aimed to minimize total cost using multi-criteria optimization. Aiming to maximize the surplus of personnel was a secondary focus. A brand new genetic algorithm was proposed containing a heuristic that's role was to provide flexibility. The genetic algorithm was written in C language. The model was compared against computational experiments and resulted in promising solutions. The authors have suggested studying stochastic programming to improve GA.

Koole, G., & Van Der Sluis, E. [4] implied in their research that service level constraint (SLA) is almost always considered in this type of problem. A study on a shift scheduling problem for call centers had been done considering global service level agreements. The method that was used in this research is multimodularity. The methodology contained multimodularity and local search, respectively. It was also stated that the local search algorithm stops when it reaches the global optimal solution. Afterwards, employee scheduling is considered together with multimodularity. The authors indicated -for small instances- that the optimal schedule was reached. In this particular study, the call center was not always in operation, and the shifts were presumed to have a similar length without breaks. The authors also implied that minimizing a multimodular function could be considered equivalent to employee scheduling. Simulated-annealing (SA) heuristic was underlined to be used in such cases where the complexity of the problem increases.

Bard, J. F. [5] worked on employee scheduling problems taken into consideration in this study. The aim of this study was to bring about an optimal scheduling for employees so as to satisfy various constraints labor rules and company policies. Employees had been categorized into several groups in regard to their daily working hours. Mixed integer linear programming was used. After conducting employee scheduling daily, a heuristic was employed to determine weekly plans. Considering the employees were multi-skilled, the authors implied that as the problem size increases, the probability of obtaining an optimal solution will decrease for United State Postal

Service (USPS) case illustrated in this study. The importance of intelligent heuristics were underlined in order to find feasible solutions.

Robbins, T. R., & Harrison, T. P. [6] proposed a study for personnel scheduling in call centers that aimed to satisfy service level agreement (SLA) in call centers using mixed integer stochastic programming method. The methodology of this study consists of three main stages. First stage is the formulation of the optimization problem in which server sizing and personnel scheduling were associated together. In the second stage, uncertainty in the arrival rates were identified. In the third stage, a comparison was made for the developed model against Erlang-C constraint. The key challenge that has been encountered was to satisfy a constant service level agreement with an uncertain arrival rate. It was also reported that the reduction in the cost of operations using this model was distinct. Additionally, it was put forward as a future work that the developed model could easily be adapted to queuing suppositions, specifically, relaxing the necessity of exponential service times.

Ingolfsson, A., Campello, F., Wu, X., & Cabral, E. [7] proposed a study on staff scheduling which mainly based on the method of creating the schedule and interpreting it in a cyclical manner. Integer programming was used and combined with randomization.

Brucker, P., Qu, R., & Burke, E. [8] did a literature review containing detailed information about mathematical models, project-centered planning and some special problems. The authors demonstrated main aspects of personnel scheduling problems such as determining whether or not the problem is NP-hard. For instance, minimizing the amount of changes in assignments were stated as NP-hard. It was also mentioned the importance of problem type, small sized or complex, in specific. According to the authors, complex problems can be solved by heuristics or metaheuristics while small-sized problems can easily be solved by linear programming (LP) solvers.

Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. [9] did a literature review study on staff scheduling problems. It was underlined that the existing literature mainly contains fixed inputs regarding agents. In personnel scheduling problems, the authors stated that considering multiple decisions either managerial or operational would be more enlightening. It was pointed out that most of the papers regarding this topic used deterministic approaches, meanwhile, the type of problem contains many aspects of stochastic features. The authors also pointed out that the current literature heavily consists of meta-heuristic and hybrid techniques.

Gabrel, V., Murat, C., & Thiele, A. [10] conducted an overview on robust optimization which reviewed and revealed most common application areas such as inventory, logistics, and finance. The importance of recent studies regarding robustifying stochastic optimization was underlined.

Lin, S. W., & Ying, K. C. [11] proposed an algorithm in this study in order to solve the shift minimization personnel task scheduling problem (SMPTSP) in three phases. The proposed algorithm was named as an iterated greedy heuristic. The first phase aimed to get an initial solution using a constructive heuristic. Afterward, advancing the initial solution was done using iterated greedy (IG) heuristic in the second phase. In the final phase, binary integer programming (BIP) technique was used and took it's initial solution from the results of IG in the second stage in order to find the optimal schedule. BIP was formulated in Gurobi (version 5.2). The outcomes were compared against VAWA (Volume + Wadelin's) algorithm and the deductions made from those outcomes revealed that the proposed algorithm is significantly robust, and better in terms of performance.

Excoffier, M., Gicquel, C., & Jouini, O. [12] did a research using a common technique for scheduling problems which is mixed integer linear programming. Call center case, in specific, aims for lowering the personnel cost and increasing quality of service, simultaneously. In addition to above, in this study, the problem was formulated as stochastic programming problem, and the model constructed was considered to be an extensive scale mixed integer linear programming. The key challenge encountered in the routing/assigning process was underlined as the uncertain feature of call arrival rates. CPLEX and C++ programs were used in this problem.

Walter, M., & Zimmermann, J. [13] conducted a research on staff scheduling which aimed to appoint each agent to more than one job, simultaneously. The assumption that there was no uncertainty about parameter values attracted attention as a limitation. In this study only optimization was used, mixed integer linear programming (MILP), in particular. It was first shown that the problem is NP-hard. Then, MILP was formulated and solved using CPLEX. At the end, a performance analysis was done. Computational traceability was the key focus/challenge. By using heuristic drop principle, the best performance was achieved. Also, a comparison against several others heuristics was implemented. The outcome of this study revealed that in terms of comparing solution quality and computation time, the drop principle was found to be performing the best. Additionally, it was stated that building robust solutions could be future work.

Taskiran, G. K., & Zhang, X. [14] did a research on personnel scheduling in call centers that aimed to investigate the strategic analysis of cross-training. Mixed-integer programming optimization method was used. The first step was to obtain the optimal mixture of the manpower. The second stage focused on determining weekly tours of manpower. The proposed two-staged method was compared against a commercial solver, Xpress MP, and performed better. As an outcome and future work, the authors pointed out that cost savings could easily be achieved through cross-training, especially with stochastic demand.

Mattia, S., Rossi, F., Servilio, M., & Smriglio, S. [15] did a study on call center staffing using two stage robust optimization which employed two-stage integer programming as an optimization technique. These two stages included scheduling and allocation, respectively. Robust optimization's importance was highlighted in terms of dealing against uncertainty.

Bodur, M., & Luedtke, J. R. [16] studied on staffing and scheduling in general. They introduced an overall approach containing three main steps which were finding an initial solution via stochastic integer programming, interpreting the solution via simulation, and then choosing the best among others, respectively. They used branch-and-cut algorithm based on Bender's decomposition. The key factors or limitations were quality and cost. A considerable cost improvement was achieved with the proposed SIP model.

Vermuyten, H., Rosa, J. N., Marques, I., Belien, J., & Barbosa-Póvoa, A. [17] proposed an integer programming solution to a scheduling problem of medical emergency service. Two heuristics were introduced, a dividing heuristic and variable neighbourhood descent search (VNDS) heuristic. Several instances were solved and these techniques were tested against each other. VNDS heuristic performed better in terms of both time and optimal solution comparing to others which were dividing heuristic and IP. Moreover, in an instance, IP solution found and coded in CPLEX took 5 hours while VNDS found the solution in only 1 hour. Additionally, for both dividing heuristic and IP solution, optimality gap occured with 2.15% and 9.23%, respectively.

Xu, Y., & Wang, X. [18] studied on call center scheduling using enhanced artificial bee colony (EABC) algorithm. They generated the initial population by using a heuristic called GISG. After generating the initial population, neighborhoods were set. The results of this study showed that this model can be used for large-scale call center scheduling problems because it outperformed simulated annealing (SA) and hybrid artificial bee colony algorithm (HABC) in terms of finding a good sub-optimal solution. Investigating more complicated neighborhood structures was highlighted to be crucially important, and was proposed as a future work.

Xu, Y., & Wang, X. [19] investigated staff scheduling problems in call centers. The aim was to determine the right number of personnel and shifts under service-level targets. In this study, integer programming (IP) and artificial bee colony algorithm was combined. Finding out initial solutions that covers all hard constraints and then using soft constraints to improve initial solutions were the two main steps in the methodology. Combined solution approach of integer programming and ABC algorithm was also compared against experiments that was coded and solved in CPLEX. It was seen that, for large scale problems, the developed algorithm fits and performs better. A future study opportunity was stated as developing Pareto solutions from multi-objective algorithms.

Boysen, N., Emde, S., & Schwerdfeger, S. [20] aimed to match supply and demand in the problem that has been studied on, which was crowdshipping. The key challenge was a guarantee of scheduled delivery. All data containing deterministic features were considered to be a limitation. Bender's decomposition optimization technique was used in order to solve this problem. First, Bender's decomposition model was formed. Afterward, a master model was formed containing information about which parcels were appointed to which employees. In order to check whether the solution solver found (master solution) was feasible or not, the integer solution was taken into consideration in the slave problem. Finally, a greedy heuristic was applied and the solutions were checked in terms of computational efficiency. An efficient exact solution was found. Several solutions were compared against each other. BBC (Bender's decomposition) performed better comparing to MIP coded in Gurobi. General pricing function could be considered as future work.

Mansini, R., Zanella, M., & Zanotti, R. [21] studied on staff scheduling problem. Mixed-integer linear programming technique was used. The types of constraints categorized were hard and soft constraints. Enhancing the proposed model could be achieved by employing several other objectives to satisfy the constraints. They claimed that the model developed finds a daily schedule quickly, in a few minutes. In addition to this, in terms of KPIs, specifically idle time and waiting time, the proposed model performed significantly good, and decreased both.

2.2 Literature Studies on Staff Scheduling Using Simulation

In this section, literature studies conducted on staff scheduling using simulation technique is presented.

Ernst, A. T., Jiang, H., Krishnamoorthy, M., & Sier, D. [22] did a simulation modeling for call centers. This study proposed a general outlook in terms of key performance indicators. According to the authors, the management challenges included costs, service quality, and employee satisfaction. The most common types of simulation models that have been used are traditional, embedded (ACD/CTI routing), and agent scheduling. The methodology continued with a proposed framework of simulation modeling in call centers. A future work was suggested as more in-depth analysis of distributions in order to reach more robust solutions.

Mehrotra, & Fama. [23] in another study, developed and proposed a stochastic model for call centers which was time-dependent. Three models were proposed for call centers regarding arrivals which were doubly stochastic Poisson model, more flexible covariance matrix, and a different type of flexibility. The properties of the stochastic model for arrivals included variance being greater than the mean, time dependence, and non-zero correlation. It was also underlined that the correlation between arrivals was not supported in nonhomogeneous Poisson process (NHPP) models. In addition, the arrival process has been obtained to be sensitive in the proposed models.

Whitt, W. [24] did a research on queuing models for call centers that attracted so much attention recently, when measuring performance. In this study, an algorithm was developed according to an approximation to the Markovian queuing model*(M/M/s/r + M(n))*. It was assumed that the call center had a Poisson arrival process, and service times were not exponential. The outcomes proved the approximation.

Feldman, Z., Mandelbaum, A., Massey, W. A., & Whitt, W. [25] did a study in the context of staffing problem aimed at time stability. Call arrivals were considered to align with the nonhomogeneous Poisson process and the algorithm used named as simulation-based iterative staffing algorithm (ISA). In the simulation part, Markovian special case were applied and resulted in delay probabilities with time stability. It was also suggested that this algorithm could be employed with *M(t)/G/s(t) + G* models.

Doomun, R., & Vunka Jungum, N. [26] studied flexible business process modeling using simulation for call centers. The author first presented the differences in the BPR and current application areas of it. Redesigning and reengineering part mainly contained integration definition modeling (IDEF), data modeling, data flow diagramming, and flowcharting. In the concept of reengineering, flexibility, and adaptive methodology were found to be crucial in terms of efficiency.

Ma, J., Kim, N., & Rothrock, L. [27] taken into account call center simulation. Discrete event simulation technique was used and a new framework called call center workforce simulation platform was introduced. The data of the problem was a real-life data acquired from Penn State University help desk. The proposed methodology contained mainly three matters which were producing scenario, implementing simulation, and performance analysis. The simulation implementation was made on ARENA program.

Titler, S. S., Dexter, F., & Epstein, R. H. [28] investigated staff scheduling for a pediatric hospital they acquired data from. They aimed to investigate the relationship between staff scheduling and productivity. The results showed that the workday would not be intense if personnel allocation would be done according to the type of procedure that is going to be conducted.

2.3 Literature Studies on Staff Scheduling Using Forecasting

In this section, literature studies conducted on staff scheduling using forecasting technique is presented.

Taylor, J. W. [29] studied on call center staffing problem using forecasting. Holt-Winters exponential smoothing was implemented as a forecasting technique. Arrival rates were aligned with Poisson and the results were interpretted for a short time and were found to be promising.

Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. [30] conducted a literature review study regarding forecasting. The importance of staffing and scheduling was underlined, specifically incoming call rates. Additionally, a case study was done using real-life data. A possible future work was suggested accordingly, which was creating a model that enables accurate modeling of incoming calls to the system that can handle more than one type of call at the same time.

Tang, X., Liao, X., Zheng, J., & Yang, X. [31] did a study regarding scheduling in a cloud data center. They forecasted the workload on the data center using the combination of wavelet neural network and multilinear regression techniques together in order to reduce energy consumption. After forecasting the workload, they proposed a heuristic algorithm called energy-efficient job scheduling algorithm (EEJS). The results illustrated that the proposed model works well in a data center which does not have a significantly high workload.

Chacón, H., Koppisetti, V., Hardage, D., Choo, K. K. R., & Rad, P. [32] studied incoming calls to call centers using ten months of data they acquired from a US insurance company. For short-term forecasting, they stated that the collected data was enough. In addition, the data appeared to be aligned with the inhomogeneous Poisson process. The study compared several forecasting techniques such as Holt-Winters, exponential smoothing, and seasonal autoregressive integrated moving averages with neural networks. Based on the results, there were not sufficiently enough data for deep learning techniques to come up with a better solution compared to classical approaches.

2.4 Literature Studies on Staff Scheduling Using Optimization and Simulation

In this section, literature studies conducted on staff scheduling using optimization and simulation techniques are presented.

Atlason, J., Epelman, M. A., & Henderson, S. G. [33] studied staffing problems using optimization and simulation for call centers. The aim of this study included minimizing staffing costs while considering service level requirements. Service levels were obtained by simulation which used common random numbers and a cutting plane approach was proposed to solve the sample average approximation problem. It was stated that combining simulation and optimization could be a powerful tool to solve these problems but considering also Lagrangian relaxation could be a potential future work. The cutting plane approach originated from integer programming. In this study, the authors stated the importance of more computational power which relates to the cutting plane approach.

Wallace, R. B., & Whitt, W. [34] investigated skill-based routing in this study aimed to minimize total staff. In the article, both simulation and optimization were used. Researchers provided a static priority routing scheme first. The methodology was further extended with a resource pooling experiment. Skill-based routing provisioning algorithm was developed. The key challenge was routing the incoming calls considering personnel requirements. In this study, a comparison was made between the agents having several skills and agents having all skills. The developed algorithm's results gave a near-optimal solution and showed that limited cross-training performs better comparing to the other scenario. Additionally, agent utilization was underlined as a key future study.

Harrison, J. M., & Zeevi, A. [35] studied stochastic fluid models which were introduced in order to solve staffing problem in call centers. To solve this problem, linear programming (LP) and Monte-Carlo simulation were used. Simulation results were obtained from a non-homogeneous Poisson process. In this study, the authors used fluid approximation to find the best solution. Further studies were also outlined to conduct research on processing activities and agent pool structures.

Aksin, Z., Armony, M., & Mehrotra, V. [36] did a literature review on call center operations. In this review study of inbound call center operations management, there were some deductions that are highly important to mention. Quality of service is an essential aspect. The authors proposed that the relationship between the waiting time and quality is relevant. For almost all of these studies regarding call center operations, the objective is to reduce the cost of operations and increase the quality of service. This objective is even more difficult to achieve if the management needs multi-skilled employees. Therefore, the importance of scheduling and rostering was underlined in this study in order to assign tasks to the right number of employees where they are needed. Some simulation models were also stated as commonly used models in call center operations such as Erlang-A, Erlang-B, and Erlang-C. On the other hand, it was also mentioned that the combination of linear programming and simulation was widely used. The relationship between efficiency and quality of service was suggested to be further analyzed.

Deslauriers, A., L'Ecuyer, P., Pichitlamken, J., Ingolfsson, A., & Avramidis, A. N. [37] studied both simulation and optimization which gave numerous insights regarding call center operations. Continuous time Markov Chain model was used which was assumed to be time-stationary. Employee scheduling was the aim of the study to obtain efficient working schedules. In this study, both inbound and outbound calls were allowed. It was stated that working schedules could be achieved for half-hour periods using either Erlang-A or Erlang-C models. Considering the nature of the problem, staffing, the aim is to minimize the cost and, therefore minimize the number of employees. For that, the authors suggested the use of binary search algorithms or integer programming could be beneficial. Results were evaluated by analyzing waiting times. The developed CTMC model was compared against real-life simulation and the results were promising with a difference of less than 1 percent in the KPIs that were measured.

Ertogral, K., & Bamuqabel, B. [38] in another study, again focused on staff scheduling in call centers. Generating computationally effective procedures in order to solve flexible agent cases was the aim of this study. Both optimization and simulation were used, integer programming and queue simulation, respectively. LINGO/PC 6.0 was used. The methodology contained determining agent requirements each hour across a week and then regulating these requirements using simulation*(M/M/s)* to build an optimization model. Numerical experiments were also included in this study regarding agents who knew Arabic and English. An efficient exact solution was found. It was stated that further study could be on analyzing the trade-off between customer service required and the total number of agents needed.

Cezik, M. T., & L'Ecuyer, P. [39] examined the landscape of call center staffing. Scholars have found that iterative cutting plain algorithms can also be beneficial. In this study, there were simulation and optimization phases. It was implemented on an integer program. The methodology contained several stages, such as simulation (Java), sample problem optimization continued with integer linear programming (CPLEX 8.1) with cut generation, respectively. At the end of the study, computational experiments were presented and the results were stated as good. The authors suggested that the more the complexity increases, the better it is to use practical heuristics.

Atlason, J., Epelman, M. A., & Henderson, S. G. [40] did a research concerning call center staffing which often requires both simulation and optimization. These techniques are sometimes gathered together in order to get a better algorithm or method for the problem of interest. An inbound call center was analyzed in this study, which aimed for cost minimization under service level constraints. In different time intervals, the lack of staffing levels in literature was pointed out. Therefore, to solve the sample average approximation problem, the authors used simulation-based analytic center-cutting plane method (SACCPM). Service level functions were discrete. The results showed that this model could be employed for real-life instances successfully, especially when the performance of it against other heuristics are considered. For further study, imitating the performance of finite difference method could be a promising study when pseudo gradients are considered.

Chevalier, P., & Van denSchrieck, J. C. [41] did a research that differs from others from the point of operating environment. Most of the studies in the literature are operating in B2C environment. On the other hand, this study focuses on B2B. Small deductions can be done from the information provided regarding the call center. Firstly, employees with multi-skills were taken into account in a small call center. Other researches have plenty of aspects in common with this study such as aiming for cost minimization (using combinatorial optimization, branch-and-bound algorithm) considering service level constraints. Apart from most, cross-training was enabled. Poisson process was assumed to be aligned with call arrival process. Based on the Erlang-B formula, Hyward approximation has been embraced. The results for the objective of cost minimization were found to be declining.

Van Dijk, N. M., & van der Sluis, E. [42] used both optimization and simulation which are widely used tools for most of the problems encountered in real life. In addition, they require decisions to make inherent with the type of the problem. This study mainly focused on the decision of which was executing pooling or not. Moreover, several examples were given regarding different application areas of these methods including call centers, hospitals, airports, flight catering, and assembly lines. In specific, call center case was evaluated in an experimental example in terms of mean waiting times and found to be inefficient when different calls were pooled. However, the case in which one-way overflow happen turned out to be better comparing to both pooled and unpooled cases.

Castillo, I., Joro, T., & Li, Y. Y. [43] in another study, did workforce scheduling using both optimization and simulation, three-step opportunity positioning heuristic, and discrete event simulation, respectively. The traditional approach was first mentioned as a road map that contains the matters of forecasting, staff requirements, scheduling, and control. Then the road map of the proposed paradigm was given. Evaluation of the model was made using the discrete-event simulation technique. It was also mentioned that GA and SA could also be used in this problem.

Avramidis, A. N., Chan, W., Gendreau, M., L'ecuyer, P., & Pisacane, O. [44] taken into account the problem of agent scheduling. A simulation-based optimization approach was used in this study. The technique used in optimization was integer linear programming. The methodology consisted of formulating scheduling, staffing problems, and cutting plane algorithms and then implementing neighborhood search as a metaheuristic approach. It was stated that this approach provided better solutions comparing to the previous studies on the same problem. Optimizing the scheduling and the routing of calls at the same time was considered as a future work.

Bertsimas, D., & Doan, X. V. [45] did another study regarding call centers containing a fluid model and similarly used both optimization and simulation, linear programming and discrete event simulation, respectively. Abandonments were allowed and the goal was to determine optimal staffing. Compared to data-centered model, this one was found to be more workable.

Dietz, D. C. [46] conducted a researched on different methods that are being acquired for staff scheduling problems for call centers. In order to obtain the appropriate staffing levels, the Markovian queuing model was used. Agent tours were found using the quadratic programming technique. The developed model had been successfully implemented to several service centers and resulted in a decline approximately around %15-20.

Liao, S., van Delft, C., & Vial, J. P. [47] conducted a study on personnel scheduling for call centers. Total wage costs were the main focus to be decreased under the limitation of service level. In this study, both optimization and simulation methods were used, stochastic programming and Monte-Carlo simulation in particular, respectively. The distributionally robust optimization model that was developed in this study, joined together with stochastic programming in order to decrease the total wage costs. *M/M/N(Erlang-C)* model was used to figure out waiting time of calls. The key challenge was determining the shifts without breaks. In addition, the developed stochastic programming model was also compared against the distributionally robust model which disposed of the infeasibility problem the stochastic programming model had. The authors underlined that a possible future study could be adapting heuristics that may be used for large-scale problems.

Rohleder, T., Bailey, B., Crum, B., Faber, T., Johnson, B., Montgomery, L., & Pringnitz, R. [48] conducted a case study for a call center in a health clinic that aimed to maximize service quality using both optimization and simulation techniques. These were implemented in Arena program. Optquest solver was used for optimization which employed a tabu-seacrh algorithm. On the other hand, discrete event simulation was employed as a simulation technique. Performance evaluation was done by average answering speed (ASA). The case study's results showed that, as call volumes went higher from 2009 to 2010 (approximately %12), there was a notable improvement in terms of ASA in the year 2010 compared to 2009.

Chan, W., Koole, G., & l'Ecuyer, P. [49] conducted a study in which employees were grouped for different types of incoming calls considering waiting time in queue and idle times. The objective was to optimize the coefficients of a combined policy which aimed to match the customer in the queue that had waited most with the employee in operations that had the most idle time considering SLA. The technique used was a simulation-based stochastic optimization heuristic. A modified genetic algorithm was employed and the constraints did not contain any hard constraints, instead, penalty costs were used.

Peng, Y., Qu, X., & Shi, J. [50] did scheduling which is widely used not only for call centers but also for health clinics. In this study, scheduling was conducted on walk-in patients in an open-access clinic. Both simulation and optimization was used, specifically discrete event simulation and genetic algorithm, respectively. According to the results, finding the optimal schedule depends on several indicators such as attendance rate and demand.

Defraeye, M., & Van Nieuwenhuyse, I. [51] conducted a literature review on personnel scheduling which clearly demonstrated a core classification for both simulation and optimization techniques with nonstationary demand. The study was separated into two groups. The first group contained techniques that are mostly used for performance analysis and the second group contained optimization techniques. The widely used simulation technique was highlighted as discrete event simulation. From another deduction in this study, it was stated that the fluid models are used in the systems that do have a deterministic nature. On the other hand, two-stage approaches in the optimization group which contain stochastic features were underlined to be hard to integrate into the constructed model.

Aktekin, T., & Ekin, T. [52] did a study on call center staffing. The queue of the simulation was designed according to the Erlang-A queuing system. Augmented probability simulation technique (APS) was used which had the transition from optimization to simulation inherent. Stochastic programming was the basis of the model. Incoming call rates and service times were uncertain and gamma-distributed. Random numbers were used in the mathematical model of this problem which separated the method from others in a similar research context.

Andersen, A. R., Nielsen, B. F., Reinhardt, L. B., & Stidsen, T. R. [53] did a comprehensive study on optimizing staff in an emergency department. They aimed to adequately allocate staff in order to minimize the total personnel needed in the emergency department. Incoming patients to the service were modeled according to the continuous time Markov chain (CTMC). Additionally, a recursive bound adaptation metaheuristic approach was proposed to solve the problem. The mathematical model developed first was nonlinear, however, some of the equations were also rearranged as if it was an integer linear programming model (IBM ILOG CPLEX). The result of the proposed solution approach was acceptable in terms of waiting time in the queue, because the regarding constraints were fulfilled.

2.5 Literature Studies on Staff Scheduling Using Simulation and Forecasting

In this section, literature studies conducted on staff scheduling using simulation and forecasting techniques are presented.

Shen, H., & Huang, J. Z. [54] did a study on call center workforce management. Researchers have also been using forecasting and simulation (Erlang-C) together. Inhomogeneous poisson processes have been discussed in the context of forecasting. In a queuing model, Poisson assumptions have been made -in which rate functions were roughly steady for short time intervals- for call arrival volumes. Dimension reduction has also been employed for the Poisson process. This study aimed to continuously revise current predictions. The precision of the call center staffing was found to be improved with the technique that was applied.

Ding, S., Koole, G., & van der Mei, R. D. [55] made a prediction of demand in call centers that attracted the attention of researchers. The goal was to come up with the minimum error in the forecasting process. The authors differentiated incoming inbound calls as fresh ones redials, and reconnects. It was pointed out that without taking redials and reconnects under consideration, the volume of incoming inbound calls would not be forecasted accurately. Real-life data was used for comparing both simulation and forecasting results. It was realized that the weighted absolute percentage error result was better, with approximately 3 percent improvement. The techniques that were employed to solve this problem were discrete-even simulation and autoregressive integrated moving average (ARIMA).

2.6 Literature Studies on Staff Scheduling Using Optimization, Simulation and Forecasting

In this section, literature studies conducted on staff scheduling using optimization, simulation, and forecasting techniques are presented.

Avramidis, A. N., Deslauriers, A., & L'Ecuyer, P. [56] did a review study on employee scheduling. It illustrates solution methods commonly used in the literature that have used optimization, simulation, and forecasting. The author mainly underlined five different solution methods which are demand modeling, artificial intelligence approaches, constrained programming, metaheuristics, and mathematical programming approaches. Different examples of all the above were also proposed. For instance, AI approach contains Fuzzy Set Theory because the problem itself necessitates different decision-making techniques. Preprocessing is a commonly used technique for many different application areas. It was recommended to use constrained programming in order to preprocess data so as to make it simpler. In this context, it was stated that the robustness of metaheuristics makes them easily deal with complex problems. Furthermore, it was illustrated that the combination of metaheuristics such as SA and GA tend to solve the problem of interest successfully. Numerous methods were underlined for both simulation and forecasting. Exponential smoothing and regression were given as an example for forecasting, and Poisson and Erlang for simulation.

Parisio, A., & Jones, C. N. [57] conducted another study (case study) on employee scheduling used techniques of optimization, simulation, and forecasting all together. A model was introduced to solve employee scheduling problems in retail outlets that employed two-stage stochastic programming. Mixed-integer linear programming was also used. In terms of forecasting, least squares support vector regressor was employed. The problem was modeled in Matlab and MILP was solved using CPLEX 12.0. According to the results, the solution found, performed approximately 6 percent better comparing to the deterministic scheduling package used for performance evaluation.

Ta, T. A., Chan, W., Bastin, F., & L'Ecuyer, P. [58] aimed to study a simulation-based decomposition method to solve sample average approximation. In this study, simulation, optimization, and forecasting were used. Independent simulations, stochastic optimization, and forecasting -for initial staffing- used in specific. The methodology was composed of three main stages. The first stage was to generate an initial staffing decision. Afterward, stochastic optimization and simulation were introduced. Finally, SAA was combined with the decomposition method. Focusing on a cutting plane approach to deal with SAA was the key focus. On the other hand, the key finding was obtained to be the adaptability of this model from staffing to scheduling. The model was compared against a direct approach that did not use decomposition. In terms of large-scale problems, the decomposition method outperformed others when approximating the MIP problem.

Table 2.1 presents a classification of literature studies on staff scheduling. In the Appendix section, there are five more tables which are the continuation of Table 2.1. In the Appendix section, the continuation of Table 2.1 is represented as Table A.1.

Table 2.1 : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Thompson, G. M. [1]	X			Simulated Annealing Heuristic		
Mason, A. J., Ryan, D. M.,& Panton, D. M.	X	X		Integer Programming	Queue Simulation	
Easton, F. F., & Mansour, N. [2]	X			Distributed Genetic Algorithm		
Cai, X., & Li, K. N. [3]	X			Multi-Criteria Optimization & A New Genetic Algorithm		
Avramidis, A. N., Deslauriers, A., & L'Ecuyer, P. [56]		X			Discrete-Event Simulation	

Table 2.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Atlason, J., Epelman, M. A., & Henderson, S. G. [40]	X	X		Cutting Plane & Sample Average Approximation		
Bard, J. F. [5]	X			Mixed-Integer Linear Programming		
Wallace, R. B., & Whitt, W. [34]		X			Resource Pooling Experiment	
Whitt, W. [24]		X			Markovian Queuing Model - Erlang-A	
Harrison, J. M., & Zeevi, A. [35]	X	X		Linear Programming	Monte-Carlo Simulation	
Deslauriers, A., L'Ecuyer, P., Pichitlamken, J., Ingolfsson, A., & Avramidis, A. N. [37]	X	X		Integer Programming	Continious Time Markov Chain(Erlang-C & Erlang-A)	

Table 2.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Ertogral, K., & Bamuqabel, B. [38]	X	X		Integer Programming	Queue Simulation(M/M/s)	
Cezik, M. T., & L'Ecuyer, P. [39]	X	X		Linear Programming		
Feldman, Z., Mandelbaum, A., Massey, W. A., & Whitt, W. [25]		X			Simulation-Based Iterative Staffing Algorithm	
Atlason, J., Epelman, M. A., & Henderson, S. G. [40]	X	X		Simulation-Based Analytic Center Cutting Plane Method	Simulation-Based Analytic Center Cutting Plane Method	
Chevalier, P., & Van denSchrieck, J. C. [41]	X			Branch-and-Bound ALgorithm		
van Dijk, N. M., & van der Sluis, E. [42]		X			Discrete-Event Simulation	

Table 2.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Castillo, I., Joro, T., & Li, Y. Y. [43]		X			Discrete-Event Simulation	
Robbins, T. R., & Harrison, T. P. [6]	X			Mixed-Integer Two Stage Stochastic Programming		
Ingolfsson, A., Campello, F., Wu, X., & Cabral, E. [7]	X			Integer Programming		
Avramidis, A. N., Chan, W., Gendreau, M., L'ecuyer, P., & Pisacane, O. [44]	X			Integer Programming		
Bertsimas, D., & Doan, X. V. [45]	X	X		Linear Programming	Discrete-Event Simulation	
Brucker, P., Qu, R., & Burke, E. [8]	X			Linear Programming & Metaheuristics		

3. METHODOLOGY

Simulation is a tool that describes and imitates the features of a system of interest. It is generally done by using computer software aimed at simulation. The system of interest refers to a variety of organizations that are operating or may operate in different industries.

The simplest logic behind simulation is to see how the system of interest would react if in some case at any point in time, we would have changed a feature of a process. Because the focal point behind the simulation model is trying to imitate the existing system as is. This model is structured as a collection of presumptions regarding how the system functions. These presumptions are articulated through mathematical, logical, and symbolic connections among the elements within the system.

Systems that are frequently being modeled could also be given examples as roughly manufacturing plants and facilities, banks, distribution networks, hospitals, fast food restaurants, etc. The ultimate goal of the simulation is to measure the change in the key performance indicators and make some improvements over different trials in a row.

Moving to whether simulation is an appropriate tool or not, the core concepts and features need to be understood in terms of simulation. As Banks stated, one can make valuable deductions of the significance of different variables and how they relate to each other by altering the inputs of a simulation and analyzing the corresponding outputs. Banks also pointed out that, for preparation and anticipation of potential outcomes, simulation stands out for testing new designs or policies before implementation [59].

On the other hand, there are some concepts implying not to use simulation. For instance, if a mathematical model could be formed and solved for a problem of interest, simulation is not needed. Additionally, simulation should not be used, if it is easier to perform direct experiments [59].

Knowing what pathway to take is crucial for any organization regarding any means of problems in different levels of analysis. That is the reason why academic simulation research could also be combined with decision-making techniques because of the interrelationship between them. This will also be mentioned in section 6 where future work will be discussed. Understanding the advantages of simulation is also interrelated with understanding the appropriateness of simulation mentioned before. Without spending any funds for their purchasing, brand-new hardware designs, physical layouts, and transportation systems can be used in a trial [59].

Measuring time and the relationship among different features of a simulation system is beneficial for industries and organizations. However, not only these but also bottleneck analysis can be performed in a computer-based simulation program. Bottleneck analysis helps identify areas where work-in-progress, information flow, materials, etc., are experiencing significant delays [59].

3.1 Classification of the Models

In this section, the classfification of simulation models will be briefly explained.

- Mathematical/Physical Model
- Static Model
- Dynamic Model
- Deterministic Model
- Stochastic Model
- Discrete Model
- Continious Model

3.1.1 Mathematical/physical models

Physical models, as the name indicates, suggest the physical copy of the system, sort of like building a mock-up for a system of interest. But these models do not have to be mocked, on the contrary, they could be the same physical model that represents the features of the system and gives insights into the performance. Kelton et al. stated that the initial association of the term often involves a physical representation or a scaled model of the system, occasionally referred to as an iconic model [60].

Physical models could sometimes be used just to see how the preformed system is operating. It is similar to the case where the largest fast food restaurant chains have their brand as a restaurant in their headquarters in order to see how the system performs. There are some other use cases as well such as conducting a simulation for or in a risky environment. For example, training a pilot requires great simulators. In addition to this, simulating different processes in a nuclear power plant also requires simulated control rooms.

3.1.2 Mathematical models

Regarding how the system operates or will operate, this type of model consists of a collection of estimations and presumptions, both in terms of structure and quantity [60]. The mathematical models generally consist of a number of assumptions and equations that are formed to stand for the system. Mathematical models are dependent on the complexity of the problem itself. Depending on the system, forming a mathematical model might be either simple or complex. Most of the real-life simulation problems fall into complex categories which also necessitates dealing with complex algorithms.

3.1.3 Static & dynamic models

Static model stands for and symbolizes a system at a specific moment in time continuum, also known as Monte-Carlo simulation. The main tools or components of a simulation system are processes. These specific moments in time are taken into account where most simulation processes matter. On the contrary, as the name indicates, dynamic models illustrate systems as it is. The systems showing the change over time are all dynamic systems.

3.1.4 Deterministic & stochastic models

Randomness is an important characteristic determining whether the system being analyzed and modeled is deterministic or stochastic. Deterministic models possess a defined set of inputs that will yield a distinct set of outputs [59]. For instance, most of the time, scheduling an appointment would be considered stochastic. Because the customers or people who took that appointment would not be able to come to the scheduled meeting in time, in most of the cases. This is where stochastic systems intervene. Random inputs are yielding random outputs. So, basically, the problems that are defined and solved by stochastic simulation techniques could also be considered as yielding more realistic answers rather than deterministic ones. However, these are sufficiently hard to model where different contingencies are also taken into consideration.

3.1.5 Discrete & continuous models

Discrete and continuous models have a significant, unique difference. A continuous model represents a system where variables of interest change dynamically over time, in a continuous manner which means there is a continuous linkage between the end of the previous time interval and the start of the next time interval.

On the other hand, discrete systems generally suggest that discrete time points have discrete variables which can also be put forward as instant change. In discrete and different time intervals, there may be different numbers of customers. Therefore, the linkage between the previous time interval and the next one is not continuous where the gap in the number of customers in different time intervals is taken into account.

Additionally, there are such systems that contain the characteristics of these two together. In a single model, it's possible to incorporate aspects of both discrete and continuous change, known as mixed continuous-discrete models. For instance, consider a refinery where the pressure inside vessels changes continuously while shutdowns happen discretely [60].

3.2 Types of Simulation Models

In this section, some simulation models that are highly used are going to be explained in detail which are Monte Carlo, Discrete-Event, and Continuous simulations.

3.2.1 Monte-Carlo simulation

Direct analytical methods are hard to implement when dealing with simulation operations. Simulation models and programs mostly conduct iterations and generate results based on those, to finally compare one another in different circumstances. Monte-Carlo simulation, which also falls into the static type of simulation category, is a unique type of computational algorithm that uses recursive random sampling to obtain and calculate the outcomes. It is usually employed for two main purposes which are the numerical integration of functions of different simulation problem models, and risk assessment [61].

Numerical integration of mathematical functions, especially complex ones, is easily handled by using Monte-Carlo simulation which employees recursive sampling that finally yields the desired outcome. In this context, it is worth pointing out that there is no single Monte-Carlo technique, in contrast, there are many of those. The main intersection point of these is random number generation.

Random number generation, in other words in this context, Monte-Carlo sampling is the crucial point that needs to be explained in detail. Monte-Carlo algorithms use random number generation (RNG). Because of the recursive replications through algorithms, randomly generated algorithms are not genuinely random, but rather random in statistical context when signifying pseudo RNG as a side note [62].

In subsection 3.1.3, we mentioned that the Monte-Carlo simulation falls into the static type of simulation category. The reason behind this is that the state variables do not evolve while the simulation process continues in static systems, however, in dynamic systems, state variables change and evolve randomly. This is the reason why dynamic simulation covers discrete event simulation.

3.2.2 Discrete-Event simulation

As mentioned before in subsection 3.1.5, in order for a system to be categorized, and named discrete, and discrete event simulation, respectively, the state variable needs to change over discrete time intervals.

The general idea and logical formulation and creation of discrete event simulation lies under a clock and event list. Each event in the system has its planned sequence of happening, which all bring about the event list. A major feature of DES is that if an event does not occur, the state variables do not change. The general pseudo code of the DES algorithm starts with settling the simulation clock to zero and generating initial events. Afterward, searching for the most imminent event, and executing the simulation for a sufficient number of replications [62].

Byoung Choi also described discrete-event simulation with the following example. He considered a system that has both inner and outer components that are; buffer, machine, and outside World, respectively. In this system, events were considered to be arriving, loading, and unloading. Therefore, as shown in Figure 3.1 when a job comes to the system, the counter counts as having one more job in the buffer immediately. Similarly, after the processing is done and the job is unloaded, the machine is set to idle instead of busy. These changes are illustrated in Figure 3.2, which shows the trajectory over time in the system that occur in a discrete set of points for the simulation time [61].

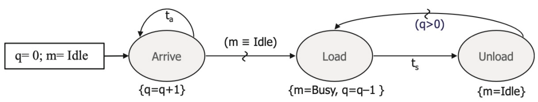

Figure 3.1 : Event graph describing the system dynamics of the single server system [61].

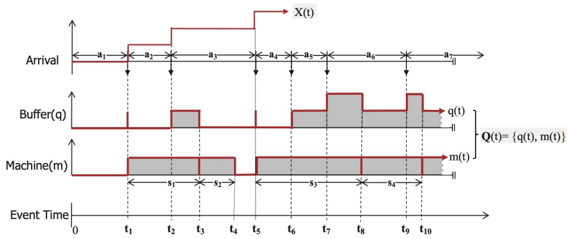

Figure 3.2 : Simulation model trajectory of the single server system [61].

A similar example was given by Altiok, considering again a similar system. The progression of the system state over time, denoted as $S(\mathrm{t})$ is represented as a step function. Discrete events trigger abrupt discontinuities in this function, leading to state transitions within the system at specific time points [62].

3.2.3 Continuous simulation

Accordance with a set of mathematical equations that generally includes differential equations capturing system reaction for the simulation time in a dynamic physical system is called continuous system simulation, which in other words, continuous numerical assessment of a dynamic, physical system [61].

As Byoung Choi illustrated in his book, thinking of Q_{t} and X_{t} standing for state and input trajectory vectors of the system, the numerical assessment of the state transition function

$$\frac{d}{dt}Q_{(t)} = A \times Q_{(t)} + B \times X_{(t)} \tag{3.1}$$

where A and B are coefficient matrices, yields to continuous simulation system [61].

Time discontinuity is not allowed in this type of simulation. This can also be further expanded by proposing more examples in this context. Imagining a dam with water behind it would be an appropriate example to describe the continuous systems in a better way. As time passes, the water level behind a dam changes according to several events such as evaporation, water release from spillways, etc. However, taking the integral of the water level change function would yield a continuous curve, and this fact matches with the limitations of continuity theory in mathematics while avoiding temporal discretization which occurs in discrete systems such that the limits of both the right and left side of function in a point of interest in time differs. Different limits would break the continuity theory laws.

As shown in Figure 3.3, taking a direct cut from any discontinuity point above, say t_1, drawing a vertical line would clearly show that approaching t_1 from both sides, left and right, would yield different limit values.

$$\lim_{t \to t_1^-} t_1 \neq \lim_{t \to t_1^+} t_1 \tag{3.2}$$

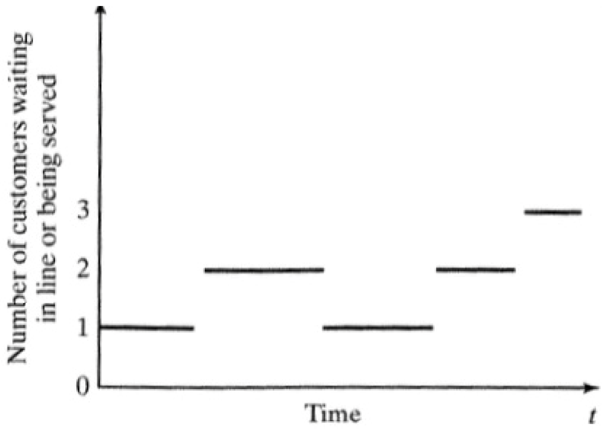

Figure 3.3 : Discrete Systems
[59].

On the other hand, the right and left-hand side limits of any specific point in time in the Figure 3.4 yield to the same value, which also satisfies the continuity theory.

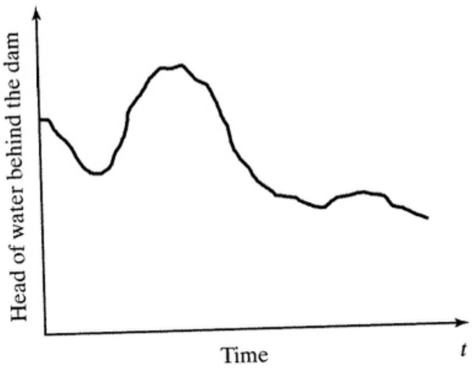

Figure 3.4 : Continuous Systems
[59].

$$\lim_{t \to t_1^-} t_1 = \lim_{t \to t_1^+} t_1 \tag{3.3}$$

In summary, instant changes in state variables of a system would be a discrete system rather than continuous because temporal discretization is what happens in those systems.

3.3 Queuing Theory

Queues are the essential visual expressions of real-life simulation studies regarding mostly manufacturing and service systems. The Queueing Theory can be defined as the mathematical study that analyzes every aspect of a queue of interest in a specific system. The queues are classified according to Kendall's notation.

Kendall's notation plays a crucial role in categorizing the queues according to their different type of characteristics. The long version of the formulation can be denoted as *A/S/c/K/N/D*. The first element *A* stands for the arrival process, in other words, the probabilistic behavior of interarrival times. In most of the cases, the arrival process is denoted as *M*, which indicates that the interarrival times are Markovian which also suggests that these interarrival times are memoryless.

The service time distribution, *S*, as the name implies, stands for the time of service of a customer. Commonly, this feature again aligns with the Markovian case, however, it can also be deterministic or other types. The number of servers is represented as c while the number of places in the queue, calling population, and queue's discipline are denoted as *K*, *N*, and *D* respectively.

For example, taking a hypothetical *M/D/k/C* system into account would give various information regarding the system's behavior. *M* illustrates that the interarrival times are independent and identically distributed exponential arrival times. On the other hand, *D* implies that the service times are deterministic while *k* and *C* stand for the number of servers and capacity, respectively.

3.3.1 Single server queues (*M/M/1*)

The most popular and known notation is *M/M/1* queue which illustrates three characteristics of a queue. The first *M* stands for the process of arrivals while the second *M* illustrates the service time distribution. The last object of Kendall's notation shows how many servers exist in the system which in this example, stands for a single server queue. This can be further explained in the Figure 3.5.

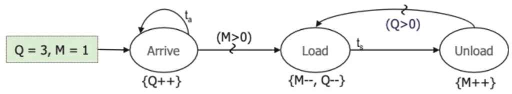

Figure 3.5 : Event graph of the single server system [61].

Figure 3.5 represents a single server queue system. Q and M stand for the state variables of the machine and buffer. Buffer is the place where jobs are assumed to be waiting because the machine is not idle. The above single server queue works as follows. Incoming jobs are being taken to the system with an interarrival time of ta. If the machine is processing a job when another comes, the job is stored in the buffer until the machine comes to an idle state. After the job is loaded, it is then processed by the machine for a processing time ts, and then being unloaded.

3.3.2 Multi server queues (*M/M/c*)

Another type of server that needs to be illustrated is a multi-server system. The main difference between single and multi-server systems is that in multi-server systems there are two or more service facilities that provide identical service in parallel. In the below figure, a flexible multi-server system with fluctuating arrival rates is given which denotes an event graph containing fluctuating arrival rates and more than two servers at time *t*. These types of queues are also known as Erlang-C queues. Figure 3.6 illustrates a flexible multi-server system with flactuating arrival rates.

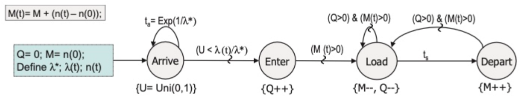

Figure 3.6 : Event graph of flexible multi-server system with fluctuating arrival rates [61].

3.3.3 Service disciplines

Based on the customer arrival pattern, service procedure, server quantity, service order, and buffer size, a service facilities' queueing system can be defined mathematically [62]. In the arrival process in the simulation system, interarrival times generally are independent and identically distributed. In queues, one of the most important characteristics is queueing discipline which stands for according to which rule the jobs are being queued up. There are numerous rules commonly used in terms of queueing discipline, but the most commonly used ones are FIFO (first in, first out), LIFO (last in, first out), SIRO (service in random order), RR (round robin), PS (priority service).

For instance, implementing a simulation with the FIFO model enables jobs to enter the queue according to the time they came in terms of rank. LIFO on the other hand, does the exact opposite of FIFO which takes the most recent arrival first into the queue.

To gauge the current situation and offer direction on the types of simulations that might be suitable for the next phase of the project, numerous people believe that queueing theory can be beneficial as an initial approximation [60].

The following Figure 3.7 represents a general simulation methodology. Although there are different methodologies, they are more or less the same. At the end of section 3, a methodology will also be presented that represents the steps that were followed to solve this staff scheduling problem.

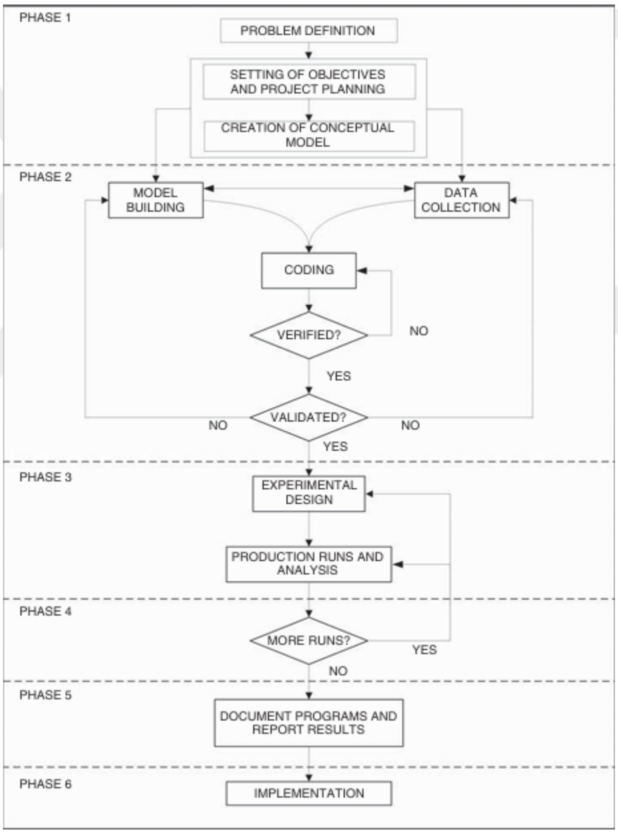

Figure 3.7 : General methodology of simulation [63].

3.4 Input Analysis

Input analysis has crucial importance when dealing with simulation problems to obtain reliable results. Input analysis is generally done by four main stages which are data collection, data analysis, time series data modeling, and goodness of fit testing. Input analysis is mainly done to observe the data characteristics, and behavior, and find a best-fit distribution that will represent the system or system part. Because in the simulation application, one of the main features is using random numbers. Those random numbers are going to be generated accordingly with the statistical distribution type that they correlate most.

In this context, it is beneficial to give some main statistical distribution types that are widely used. These distributions will not be analyzed in detail but will provide sufficient understanding to the extent to understand the problem of interest and solve it using these.

3.4.1 Probability distributions frequently used in simulation

Some main probability distribution types frequently used are represented in this subsection.

- Binomial Distribution

  1. Binomial distribution falls into the category of discrete distributions. Random number x denotes the number of successes in n trials. The trials are independent with probability p.

- Negative Binomial Distribution

  1. Similarly to binomial, negative binomial distribution models the number of trials until k success.

- Poisson Distribution

  1. Poisson distribution quantifies the occurrence of discrete events within a specified temporal or spatial constraint.

- Normal Distribution

  1. This model characterizes the dispersion of a procedural sequence that can be conceptualized as the aggregate of several constituent processes. For instance, it delineates the duration to construct a product, comprising the summation of individual time intervals necessary for each assembly operation [59].

- Lognormal Distribution

  1. This model delineates the distribution of a process, which can be conceived as the result of multiplying several constituent processes. For instance, it illustrates the compounding rate of return on an investment, where each period's return contributes to the cumulative product [59].

- Exponential Distribution

  1. This model examines the interval between autonomous events or a process characterized by memorylessness (where knowledge of elapsed time provides no insight into the duration until completion). For instance, it investigates the intervals between the arrivals of numerous customers, each acting independently of the others [59].

- Gamma Distribution

  1. Nonnegative random variables are modeled according to gamma distribution.

- Beta Distribution

  1. Utilized for modeling bounded random variables with fixed upper and lower limits, the beta distribution is renowned for its remarkable flexibility. It can be shifted away from zero through the addition of a constant and extended beyond the conventional range of [0,1] by multiplication with a constant, thereby expanding its potential scope [59].

- Erlang Distribution

    1. This modeling framework encompasses processes that can be conceptualized as the aggregate of multiple exponentially distributed phenomena. For instance, a computer network failure scenario involves the concurrent malfunction of a primary computer and two backup units, each governed by an exponentially distributed time-to-failure. Notably, the Erlang distribution emerges as a distinctive manifestation within the broader domain of the gamma distribution.

- Weibull Distribution

    1. Modeling the time until failure for components, such as a disk drive, exemplifies the utilization of the exponential distribution, which represents a specific instance within the broader framework of the Weibull distribution [59].

- Uniform Distribution

    1. The discrete or continuous uniform distribution epitomizes complete uncertainty, wherein every potential outcome holds equal likelihood. This distribution is frequently misapplied in scenarios devoid of available data.

- Triangular Distribution

    1. The triangular distribution is employed to represent a process in cases where only the minimum, most likely, and maximum values of the distribution are ascertainable, as demonstrated by scenarios such as determining the minimum, most likely, and maximum time required to test a product [59].

- Empirical Distribution

    1. Empirical resampling entails drawing samples directly from the collected empirical data, a method frequently employed when no suitable theoretical distribution is deemed applicable [59].

3.4.2 Data collection

Any type of problem that can be categorized as quantitative requires an adequate problem formulation. Afterward, taking problem objectives and constraints into consideration needs to be done to settle the fundamentals of the problem. Model conceptualizing and data collection stages go in parallel to one another. Previous stages help the analyst to understand the characteristics of the problem and conceptualize the system accordingly.

Data collection task has great importance in acquiring reliable and accurate results when dealing with real-life problems. If the collected data is somewhat insufficient or misleading in a way, the results of the simulation will also be misleading and inaccurate. Even when data are accessible, they are seldom documented in a format conducive to direct utilization for simulation input modeling purposes [59].

Data collection does not mean collecting every available data of the facility that is being observed. Some data might not even be related to the simulation problem at all. That is the reason why the inspector also needs to analyze the data collection process, eliminate the data that is not related to the problem, and observe the system while the data is being collected to better understand the system dynamics. Another important point that needs to be made is trying to combine homogeneous data sets.

The data that are being collected are generally measuring the starting and ending time of an individual task, or series of tasks. It is important to accurately define and observe a task's start and end time to capture and obtain results more accurately. To avoid the above circumstances, before collecting the data, planning a roadmap would be beneficial. Additionally, a pilot run for the planned roadmap would even be more beneficial and could be considered as a time-saving activity.

3.4.3 Data analysis

The data analysis process comes after the data collection process is done. During the analysis phase, it is common to calculate diverse empirical metrics from the gathered data [62]. These metrics can be mean, standard deviation, coefficient of variation, etc. Using these metrics, visualization is generally done, such as plotting a histogram regarding the data. It could also be an autocorrelation plot.

By conducting this study, the analyst is not only visualizing the data but also expressing the basic statistical characteristics of the data. In the end, the data analysis stage will help the analyst to find the best-fit distribution for different processes. The basis of the data analysis stage is to re-process the data to eliminate unnecessary ones that do not correlate in any way with the system behavior.

Some statistics generally gathered from the data are as follows:

- Mean,
- Standard deviation,
- Coefficient of variation, etc.

Using these statistics, in order to visualize the data and find the type of distribution that the data aligns with, the analyst also should plot a histogram, or several of those, depending on the data and the problem itself. For instance, the box plot method could be used to determine the outliers of the data. Accurately analyzing the data also enables the analyst to find best-fit distributions that behave just like the system itself. In this way, random numbers are going to be produced accordingly with the type of distributions determined.

3.4.4 Determination of best-fit distribution

The distribution types defined in subsection 3.4.1 were meant to be general information. The basis that stands for the methodology of solving staff scheduling problems, similar to many others, starts with data collection and then continues with data analysis.

In data analysis, some statistical features of the data are determined so as to observe how the processes behave in the system of interest. Additionally, it is important to point out that data pre-processing is also being conducted in this part such as determining the outliers in the data using some statistical techniques like the box plot method. Extracting them from the data enables the analyst to come up with more accurate results. However, in some cases, extracting every outlier from the data may also lead to inaccurate results. The reason behind this lies in the nature of the data. It may be like that particular data has some outliers in it.

Although the solution techniques or methods being used to solve the problem are not stochastic, it is crucial to keep the stochastic nature of the data in order to mimic the system to finally create a system that behaves just like the original system that was observed. For instance, in a call center case, assuming the interarrival times to the system as five minutes on average with a standard deviation of 2 minutes, should not always direct the analyst to think that ten minutes interarrival time would be an outlier. Because this behavior lies in the nature of that system, although the number of such outliers is generally very low considering the main data, could be estimated maybe as 1% of the whole data.

Moving to best-fit distribution in the next stage, the logic behind this process, as explained before, is to create a system that perfectly mimics the actual one, so as to analyze and solve the scheduling problem and propose a better solution to the previous one.

After collecting and pre-processing the real data, in order to create maybe one of the most important features which is random number generation, trying to find a statistical distribution type that the data best fits is crucial. There are many tools that can be used to conduct this either by conducting calculations by hand and testing every available statistical distribution against the data in order to figure out which one is more suitable, or performing a computer analysis by using some programs that are aided for this process such as ARENA.

ARENA simulation software will be explained in detail in the following sections. Its relation to this section is that it has a built-in tool that enables the analyst to perform the best-fit distribution analysis, named as Input Analyzer. This tool takes the initial original data as input and tests it against the statistical distribution types that are again built into it. Moreover, the Input Analyzer tool visualizes the data by plotting a histogram. Additionally, most of the statistical information that the analyst needs is calculated and shown below the histogram as a summary of the original data.

Determining best-fit distribution means that the analyst has a statistical distribution that behaves just like the original data. From now on, every timing information for processes or interarrival times can be created using the specific distribution type as an input generator to the system. Best-fit contributions may differ from process to process, or from system to system based on the characteristics of that specific system.

It is also important to point out the importance of the sample size. Understanding the impact of sample size is of paramount significance. In situations where data are scant, the likelihood of a goodness-of-fit test failing to reject any candidate distribution increases. Conversely, in scenarios characterized by ample data availability, the probability of a goodness-of-fit test rejecting all candidate distributions significantly escalates [59].

3.5 Validation & Verification

As shown in Figure 3.7 validation and verification have a significant importance. Model building and data collection are closely related and connected to the validation and verification phases. The ultimate aim of validation is to generate a model that faithfully replicates the genuine behavior of the system to such an extent that it can serve as a surrogate for the actual system in experimentation endeavors. Additionally, another aim is to enhance the credibility of the model to a level deemed acceptable, thereby fostering its adoption by managers and other decision-makers [59].

In other words, by conducting the validation phase, the analyst will be able to determine if the model built entirely and adequately represents the real system. It can simply be understood from the results of the simulation, specifically, comparing the outputs of the system modeled and the real system to see if there exist any inconsistencies.

Verification on the other hand pertains to the accurate construction of the model, focusing on aligning the conceptual model with its corresponding computer implementation. It involves assessing whether the model is implemented accurately within the computational framework and whether the input parameters and logical structure are faithfully represented [59].

Model building and data collection go in parallel with interaction to both validation and verification. This also means that the system model may evolve in time with the feedback mechanism that is illustrated in Figure 3.7. If the model is not accurately constructed, verification will not be accurate redirecting the process back to the coding stage. If the process is verified but not validated, that also means that there needs to be a backward process either toward model building data collection, or both. Model verification can also be conducted through queuing systems performance analysis. The below figure illustrates a generic simple workstation.

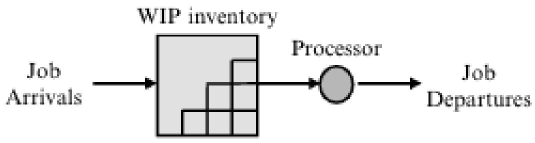

Figure 3.8 : A simple workstation [62].

The simple Workstation shown in Figure 3.8 represents a generic queue example in a manufacturing facility. The procedure starts with the incoming jobs moving to the processor. If the processor is busy, they are generally kept in a buffer (work-in-process, abbreviated as WIP).

In manufacturing facilities, workstations, and work stages are generally being modeled as a queue system in which keeping records makes arrival and service times crucially important for performance measures. In the processor, jobs are handled by queueing discipline explained previously such as FIFO (first in, first out). Performance analysis will be further broken down in detail in section 4 in which output analysis will be explained in detail. Little's formula intervenes when verifying the model. Little's formula is illustrated with the following equation.

$$N \equiv \lambda \times W \tag{3.4}$$

λ generally stands for the arrival rate, however, in this equation it is modified, meaning that lost jobs are excluded. N and W stand for the average number of jobs in the system and the average time a job spends in the system, respectively. Little's formula, renowned for its simplicity and universality, serves as a vital instrument in validating queueing simulations. Its utility extends to the detection of modeling inaccuracies or coding anomalies, particularly in cases where customers are either stranded within the system erroneously or inadvertently generated or eliminated [62].

3.6 Output Analysis

Output analysis comes in the next stage after verification and validation of the model are done. The simulation process begins with some goals at first and continues in such an aspect to reach those goals. By analyzing quantitative data resulted in after the simulation process is done, the decision maker or manager can make qualitative decisions.

This stage is in most of the methodologies, the last step. Because the model is already built, verified, validated, and run. The remaining process that needs to be handled is mainly output analysis and enabling the decision maker to draw some decisions from it. Figure 3.9 shows a general framework of simulation output analysis.

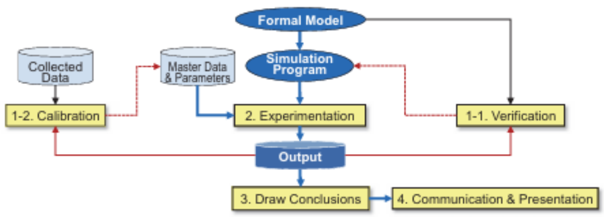

Figure 3.9 : Framework of simulation output analysis in discrete-event simulation [61].

In Figure 3.9, a framework used in output analysis is shown. In Figure 3.7, after creating the conceptual model, the framework shows a direction leading to data collection and model building in phase 2 which they both interconnected with validation and verification tasks. Similarly in another methodology, the general framework more or less overlaps.

The formal model comes after building the reference model, which also stands for conceptual modeling given in Figure 3.7. Between the reference model and the formal model, the validation task is done. On the other hand, between the formal model and output stages verification task is conducted. In the third matter in Figure 3.10, drawing conclusions generally means comparing outputs against each other for a sufficiently enough number of iterations so as to come up with a credible conclusion that could be beneficial when making qualitative managerial decisions.

Some queueing performance measures are illustrated below.

- Average number of jobs in the queue
- Average number of jobs in the system
- Average job waiting time
- Average job delay time
- Utilization rate
- Throughput

In the simulation outputs, what the analyst focuses on is mainly the value-added times that were pre-defined when the simulation model was built. As the name suggests, if the model was built and run accurately, the output statistics will enable analysts to see the performance measures listed above. For instance, by just looking at the utilization rate in every processing unit that has a service time, the analyst can easily see the bottleneck in the queue system.

In this hypothetical instance, the manager most likely would want to decrease the utilization rate and thereby also relax the bottleneck by maybe adding some machines in parallel lines or by changing the number of employees that work in the peak time of service needed.

The above example was only given to convert the hypothetical and theoretical cases to more concrete ones in order to better illustrate how the simulation process as a whole works out. Another important feature of output analysis is to initially determine the number of replications before running the simulation. If the pre-determined number of replications is not enough, most of the statistical metrics the simulation output page presents will not be sufficient. Half-widths of the important features will not be calculated due to an insufficient number of replications.

Graphical simulation outputs contain both physical and logical animations of the simulation model. The type of outputs used change accordingly with the case that is faced. Logical animations account for one of the most important and valuable outputs during the initial stage of simulation whereas physical animations are useful to catch errors.

Simulation outputs give the analyst every available outcome that was aimed to be found in the beginning. In the output analysis stage, the decision maker generally asks the analyst to find the key performance indicators that represent the main statistical information in a form like a graph.

Performing scenario analysis comes next when the ultimate goal is to improve a specific process or overall system in a way that is desired. Decreasing the waiting time in queues might be an example of a managerial goal. Moving towards that goal necessitates some changes in different scenarios defined after the main simulation for that facility is completed. Scenario analysis mainly takes the data used in the first simulation as scenario 1. Afterward, the analyst changes the different features of the simulation data so as to reach the managerial goal that is desired such as decreasing waiting time in queues.

The analyst may suggest the manager hire more employees during busy hours so as to minimize the average waiting time in queues, which also accounted for one of the most important key performance indicators. So, in scenario 2, the analyst may try to observe the outcome of scenario 2 and compare it to the first one by also pointing out the change in the money spent.

Another important characteristic of output and scenario analysis is to realize that in almost all of the scenarios, there is a trade-off. When the managerial aim is decreasing the average waiting time in queues for customer satisfaction, hiring more personnel will yield an increase in money spent. The decision maker must choose between these two, which makes this situation a trade-off.

Additionally, decreasing again the utilization rate, which in other words means decreasing the usage of that specific machine or processor, will relax the bottleneck. However, if the goal of decreasing the utilization rate is achieved by employing more machines or processors, the cost of buying those will again be a matter that the decision maker needs to decide which is also another trade-off.

These statistical outcomes or performance measures that were given above can be plotted on a graph. The beneficial part of this plot is that it visually makes the comparison of different alternative system designs more easy and accurate. These output plots represent the system behavior best, comparing others.

3.7 Simulation Tools

In the subject of simulation application, simulation programs also matter because these programs are also categorized. Depending on the project, there are several widely used simulation programs that are aimed mostly at that specific type of project. In this part, some simulation programs that are widely being used worldwide will be listed and briefly explained. The list of some simulation programs is shown below.

- Simulink

  1. Simulink is a built-in program interface that comes with Matlab. It provides various features in it for different types of simulations. It also contains some beneficial functions compared to other programs such as the ability to optimize simulation parameters. Due to its strong capability, it enables running thousands of simulations in parallel. Moreover, analyzing simulation results and deploying those are easy in its interface, and user-friendly. Various types of simulation projects can be conducted with Simulink.

     (a) Artificial Intelligence

     (b) Wireless Communication

     (c) Electrification

     (d) Control Systems

     (e) Signal Processing

     (f) Autonomous Systems and Robotics

     (g) Advanced Driver Assistance Systems

     (h) Digital Twins

- Arena

  1. Arena is a powerful software designed mainly for discrete event simulation and automation. Conceptualizing in simulation projects is easy to achieve with Arena considering the tools that are provided within the interface. Processes, workflows, and nameless modules are available for the analyst to conceptually design the project, control the parameters, and run the simulation with the desired conditions. It also allows the user to perform scenario analysis which is crucial to make in order to compare several different alternatives. Additionally, both visual and statistical outputs of the simulation project are provided in a well-designed output file. The types of simulation projects that can be handled with Arena are listed below.

     (a) Manufacturing Processes

     (b) Supply Chain Management

     (c) Healthcare Systems

     (d) Service Systems

     (e) Logistics and Transportation

     (f) Business Process Improvement

     (g) Project Management

     (h) Digital Twins

     (i) Facility Layout Design

     (j) Risk Analysis and Decision Support

- Anylogic

  1. Anylogic is another powerful simulation program that is widely used in several different types of projects.

     (a) Discrete Event Simulation

     (b) Agent-Based Modeling

     (c) System Dynamics

(d) Hybrid Simulation

(e) Digital Twins

(f) Optimization and Decision Support

- SIMUL8

1. Another powerful simulation program is SIMUL8. It is a discrete event simulation software for modeling and optimizing processes. Different types of projects that can be performed with this software.

   (a) Manufacturing and Production Processes

   (b) Supply Chain and Logistics Management

   (c) Healthcare Systems

   (d) Service Operations and Queuing Systems

   (e) Business Process Improvement and Optimization

   (f) Facility Layout Design

4. APPLICATION

In this chapter, practical application of simulation analysis is explained and discussed in detail. Firstly, problem definition is given. Afterward, data set, input analysis, process flow chart, ARENA model, and output and scenario analysis presented.

4.1 Problem Definition

Call centers play a crucial role in terms of customer satisfaction. In the cargo industry, it holds even more importance than many others, not only for B2B but also for B2C deliveries. Long waiting times in queues cause re-dials or call abandonments negatively affecting overall service delivery. The term “call center staff scheduling” mainly refers to the effective use of resources while meeting customer demands to prevent customer dissatisfaction. The aim of this thesis is to illustrate the existing state of the operating system in a cargo call center to ultimately manage staff scheduling effectively and efficiently, using the ARENA simulation program.

Within this context, related data on call center operations for different processes is collected. As it was mentioned in the previous section, methodology, the analyst does not need every data available in order to conduct simulation analysis.

The problem is first formulated in a generic way. Afterwards setting objectives and determining the overall project plan comes. The analyst needs to gather the data that will lead him to the objectives that were set and pre-defined. The suggested framework modeled in the previous methodology section is followed in this application section.

In the framework, model conceptualizing, reference model creation, and data collection stages go hand in hand, as they should be in order to have more accurate results. By creating a connection between data collection and setting objectives stages, the analyst will be able to model and solve the problem of interest in a comprehensive way.

The problem regarding the service operations of the company included effectively handling incoming calls without allowing any re-dials, or minimizing those in order to increase customer satisfaction. Additionally, long waiting times in queues decrease the level of satisfaction. Therefore, the primary objective of the managers was to increase customer satisfaction while effectively handling operations. Also, it was the company's interest to effectively allocate their employees in their working schedules.

Effectively scheduling staff will also contribute to the decision-making process on the decision of whether or not to hire employees. This is the main part where the tradeoff takes place. Aiming for effectively and efficiently scheduling staff is likely to have a return in terms of cost.

Another important study that needs to be considered and carried out is bottleneck analysis. Almost every company that mainly is in the service sector is trying to effectively handle the bottlenecks. If there are not enough employees, in an operation system that is considered also as a queue system, waiting lines will be long. In other words, the waiting times of customers will be long. Therefore, even aiming only for reducing the waiting times in queues will be beneficial for the company to satisfy the demand for customer satisfaction.

These important matters that are explained above can also be measured in a more technical and scientific way, that is, measuring key performance indicators, also called KPIs. These main KPI's are listed below.

- Average number of jobs in the queue
- Average number of jobs in the system
- Average job waiting time
- Average job delay time
- Utilization rate
- Throughput

The KPIs above not only illustrate the current situation of the system but also give the analyst insights on the scenario analysis that will be conducted afterward.

In this subsection, the framework that is created and used in order to solve the problem will be given in order for clarity. The below figure represents the overall problem-solving methodology that is employed. In this section, the steps until “Reference Model Creation” will be considered. In subsection 4.3 in which input analysis is explained, “Data Analysis” and “Determining Best-Fit Distribution” steps will be explained. Formal model creation will be explained in detail in the subsection 4.4. Afterward, a process flow chart will be given in subsection 4.4, where simulation parameters and application will be further explained. And then finally, in subsection 4.5, output analysis will be conducted, subsequently, scenario analysis will be done.

Problem formulation starts with the definition of the problem itself. Clearly defining the current problem in the service system of interest is crucially important in order to conduct a solid analysis. The main problems the firm had listed below. Additionally, Figure 4.1 presents the simulation application framework developed.

- Not being able to foresee how many employees that need to be hired.
- Not being able to schedule the staff effectively and efficiently.
- Long waiting times in queues.
- Customer dissatisfaction.
- Call abandonments (Especially in peak hours).

In order to manage the above problems listed, the firm’s objectives were collected. Creating an overall project plan, and formal model, and then finally conducting a simulation application necessitates determining the objectives first. The main objectives the firm had are listed below.

- Measuring the KPIs in order to improve the performance.
- Scheduling staff effectively.
- Ensuring customer satisfaction.

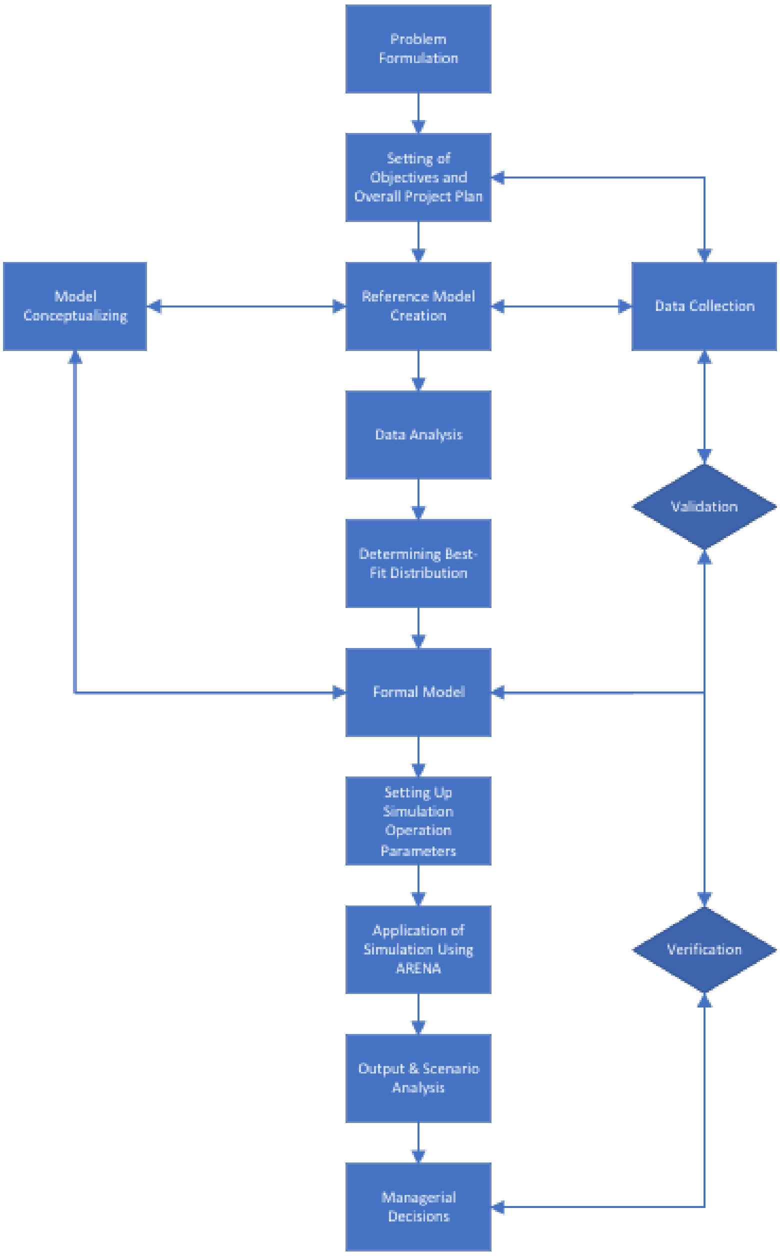

Figure 4.1 : Simulation application framework.

4.2 Data Set

Aiming to reach the objectives mentioned before, related data needed to be gathered. The data on hand included every incoming call going from 1 to 20117 for one month, March 2022. The data that the call center gathered included the followings:

- Call ID,
- The information on whether the call was assigned to the queue or not,
- The information on whether the call was answered or not,
- The information on whether the call was missed or not,
- The date of call (Given in the form of yyyy-mm-dd & hh-mm-ss),
- The date of queue (Specifying the exact time that the call was assigned to the queue),
- The date of ending the conversation,
- Connection waiting time from queue to operator,
- Call duration.

The gathered data was clarified from outliers, so we did not conduct any other pre-processing for outliers. This process is not a one-time event, but rather an iterative process. Therefore, there are some calls that could be considered outliers, however, as was mentioned in Chapter 3, methodology, disrupting the original and natural form of the data will not yield accurate results because of the stochastic nature of both the data and the job itself.

The number of these outliers is extremely low, less than 1%. Box plot methods were used to identify these outliers. In order to make the data align with the stochastic nature of this service system, we allowed some outliers to go through. Model conceptualizing and reference model creation are given in the following Figure 4.2.

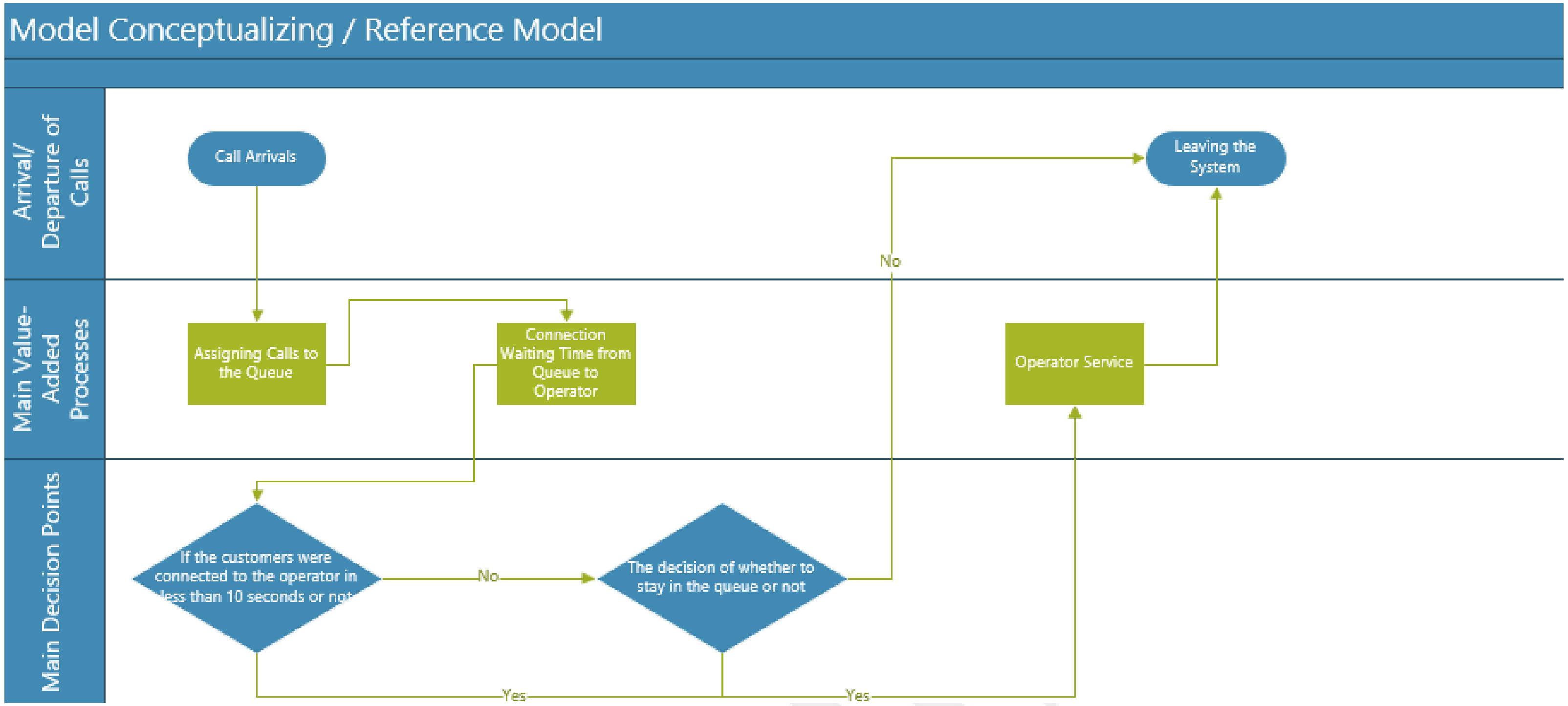

Figure 4.2 : Model conceptualizing & reference model creation.

The main stages of this simulation application were first divided and analyzed conceptually in three stages as illustrated in Figure 4.2. There were two main decision points along with three main processes that are value-added. By value-added, the key point being referred to is entering the simulation program that the time passed in that process is value-added, therefore related data needs to be measured, saved, and taken into account both in the simulation application run and output report so as to accurately obtain the relative KPI's.

The descriptive statistics of the data is shown in Table 4.1.

Table 4.1 : Descriptive statistics (in seconds).

Criterion	Call Duration	Interarrival Time	Waiting Time of Missed Calls	Duration of Assigning Calls to the Queue	Connection Waiting Time from Queue to Operator
Q1	82.00	14.00	28.00	36.00	3.00
Mean	151.64	55.43	68.84	37.01	15.09
Median	121.00	34.00	49.00	37.00	4.00
Q3	9.87	10.00	93.00	38.00	4.00
Range	1407.00	993.00	422.00	145.00	550
St.Dev	71.90	132.86	58.74	7.01	34.91

4.3 Input Analysis

Input analysis is carried out immediately after the reference model is created. This is the stage where all the necessary data is already collected. The data that was gathered from the call center included different time measurements on the incoming calls. In order to analyze the raw data, a Matlab model was created. The Matlab model prepared the input data for the Arena simulation program and is given in Appendix 1.

Important time measurements for this simulation problem included the followings:

- Call duration,
- Interarrival times,
- Waiting time for missed calls,

- Queue assignment time,
- Connection waiting time from the queue to the operator.

These durations were extracted from the Matlab model to conduct some further analysis of the data. Raw data included 27 days of incoming calls because the call center did not operate one day in every week. Additionally, every day, employees started working from 8:30 A.M. to 8:30 P.M., approximately 12 hours in different schedules.

4.3.1 Data analysis

Analyzing the data on Matlab yielded a deduction which was, every day, incoming call volumes differed from hour to hour, across the day. In other words, there were some time intervals when the call volumes were high, similarly, there were some time intervals when the call volumes were low. Therefore, differentiating the calls accordingly throughout the day would be beneficial for scheduling employees effectively and efficiently.

We differentiated the incoming calls according to their volumes. We constructed a heatmap, dividing all day hour by hour on the x-axis. Every column on different time intervals was painted from black to white, corresponding to the mean incoming call numbers, respectively. The below figure represents the heatmap of the incoming calls. These results were not only obtained from one day. For every single day in the raw data, mean incoming call numbers in different time intervals were determined. Afterward, we took the average of 27 days of these particular results to ultimately have more generic results that represent the behavior of the whole data.

These colors make it easier to schedule the staff. It is one of the main objectives to reduce the bottleneck of the service queue and allocating and scheduling staff from idle to busy hours will relax the queue and utilization rate.

An important assumption has been made for the simulation application. The raw data included every incoming call for one month. However, it is also important to take into account the call characteristics and find out which data needs to be selected for simulation analysis. Therefore, we made an assumption regarding some calls to be eliminated, which are the calls that were successfully connected to the operator but were also dropped in less than ten seconds.

The reason behind this is the assumption that the problem of the customer is not likely to be solved in less than ten seconds. This assumption could also be considered as an expert opinion and falls into the stages of model conceptualizing and reference model creation. Moreover, the assumption made to eliminate these calls also changed some KPIs in the outputs, meaning, these calls did not contribute to the missed call ratio, because these are not considered missed calls. Figure 4.3 shows the heatmap of incoming calls.

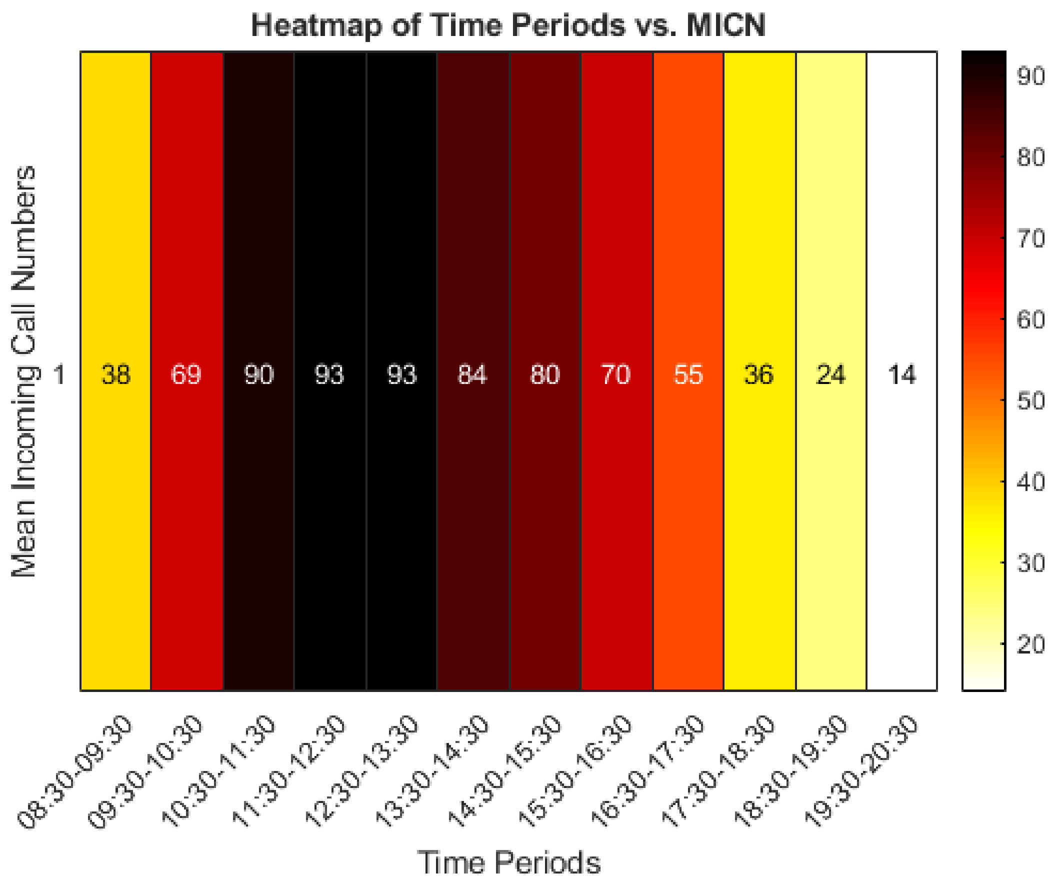

Figure 4.3 : Heatmap of time periods vs. MICN.

There were a total of 3627 calls among 20117 in which the connection waiting times were more than 10 seconds. The ones that were in the interval of the first ten seconds had already connected to the operator to be served. But our interest was to find out how many customers waited even if it took too much time.

Figure 4.4 represents the total number of customers who waited in the queue for more than 10 seconds.

By looking at Figure 4.4, it can be seen that the total number of customers who waited in the queue between 10 and 50 seconds is more than 18 hundred, approximately half of those. When the waiting time was raised to 50 and 90 seconds intervals, we deducted that it yielded almost one-fourth of 3627 customers. The remaining results were found to be 12.4, 6.3, and 6.1%, respectively.

The total number of customers who waited in the queue for more than 10 seconds was found to be 3627. In other words, the number of customers who waited in the queue but did not take service was 789. Figure 4.5 shows the frequency of lost calls by time intervals. The X-axis represents the time spent in the queue without taking any service in seconds while the y-axis shows the frequency of these calls.

An interesting deduction from both Figures 4.4 and 4.5 is that these two figures show almost the same behavior. When both of the graphs are analyzed, it is seen that in the same time intervals, comparing these two yields the conclusion that the total number of customers who waited in the queue and again the total number of customers who waited in the queue but had not been served, more or less, overlaps. These illustrations also contribute to managerial decision-making, considering the insights deducted from it.

In Figure 4.5, the customers who waited in the queue between the interval of 10 to 50 seconds and dropped the call were measured as 50.2%, whereas the frequency of lost calls between 50 and 90 seconds yielded 23.4%. The remaining columns were 13.9%, 6.7%, and 5.8%, respectively.

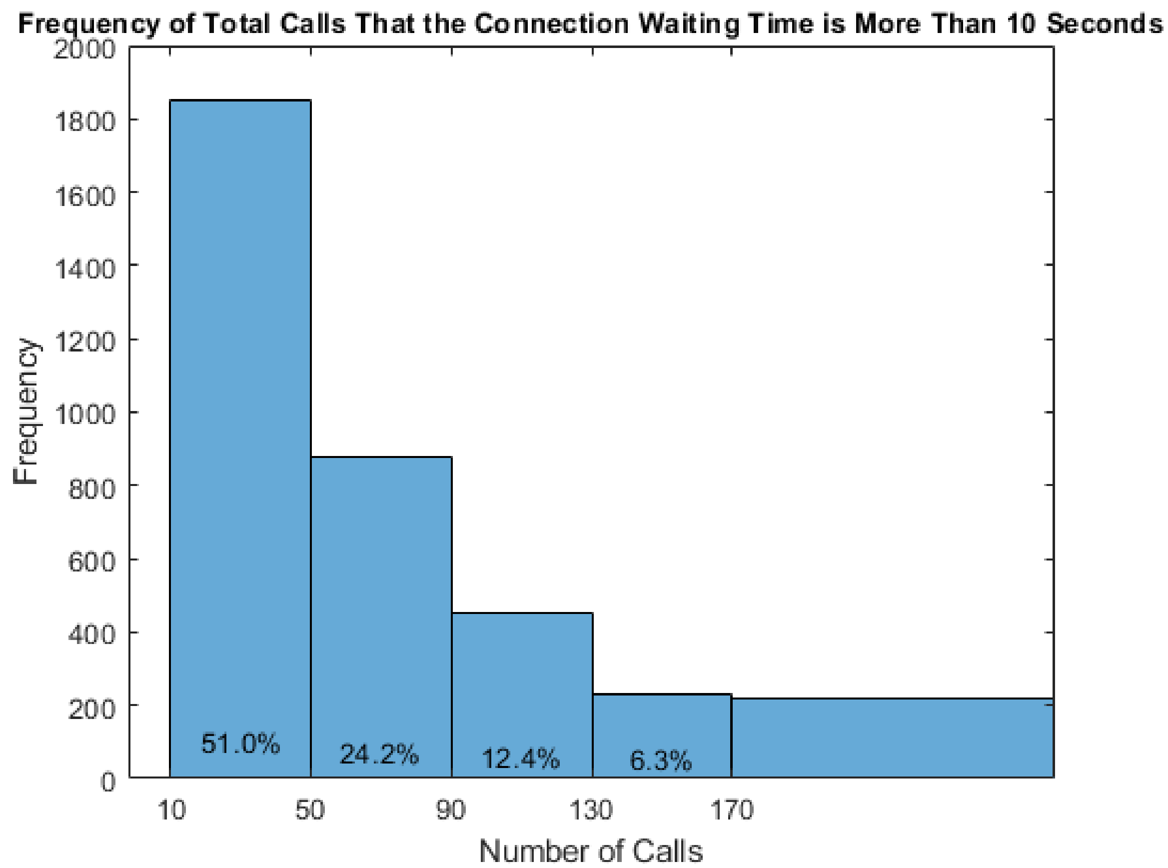

Figure 4.4 : Frequency of total calls that lasted more than 10 seconds by time interval.

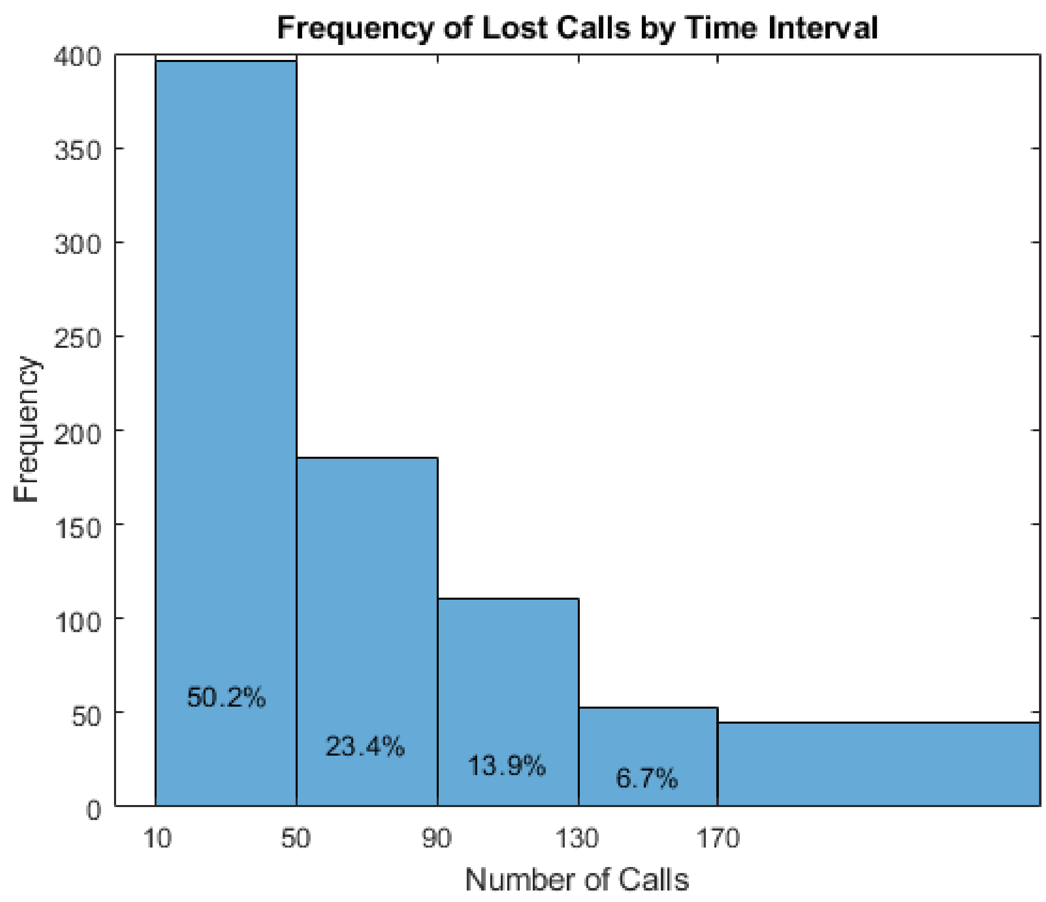

Figure 4.5 : Frequency of lost calls by time interval in seconds.

4.3.2 Determining best-fit distributions

In Figure 4.2, we did the model conceptualizing and created a reference model of simulation with its main futures such as the decisions to be made and main processes that need to be conducted. The reference model illustrates the main steps to be followed in order to conduct the simulation analysis accurately.

In addition to this, the main timings to be measured were determined with an in-depth analysis. These data again were taken from Matlab, where data analysis was done. Important data that will be received by the ARENA program are as follows:

- Call durations,
- Interarrival times,
- Waiting time of missed calls,
- The waiting time to assign the call to the queue,
- Connection waiting time from queue to operator.

4.3.2.1 Call durations

Call durations were taken from the raw data. But we made a few corrections. For instance, we extracted the number of calls that lasted less than 10 seconds from the total number of calls in which operator-customer interaction took place. The number of calls that lasted less than 10 seconds was 209. Therefore, the total number of calls where dialog took place was found to be 18704.

Thus, with the help of the ARENA input analyzer tool, we were able to find the best-fit distribution, statistical characteristics of call duration data, and its histogram which all are in Figure 4.6.

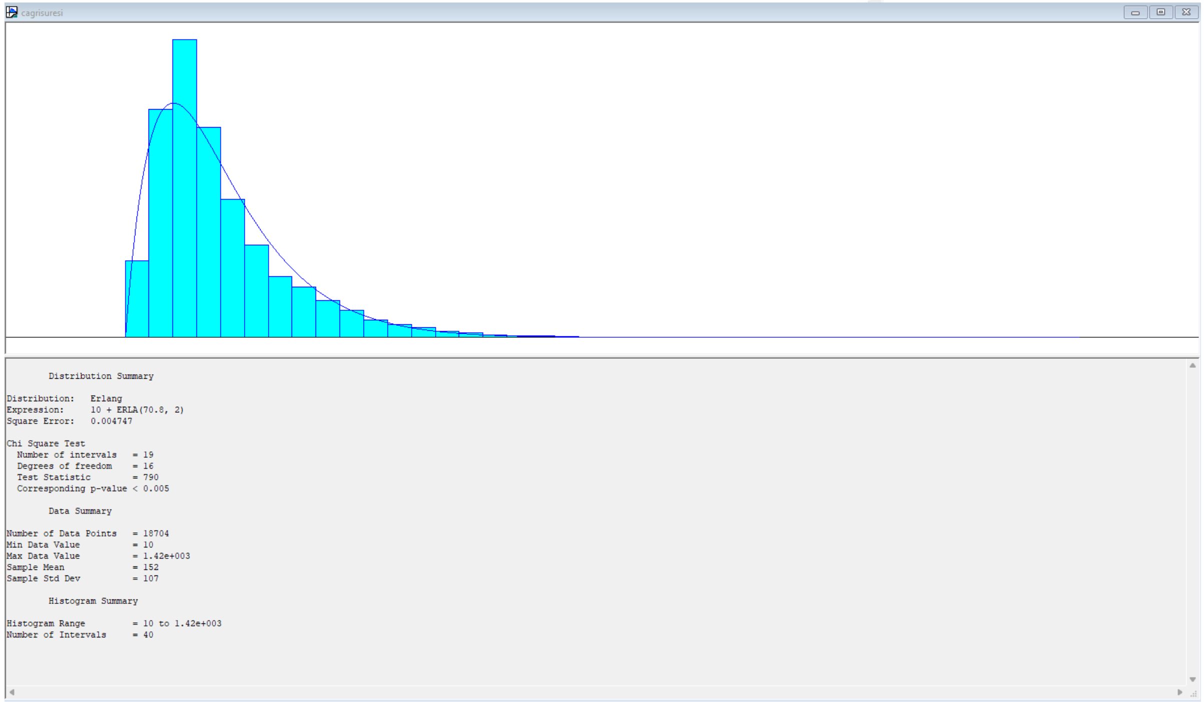

Figure 4.6 : Best-fit distribution of call durations (in seconds).

It can be seen from Figure 4.6 above that the related data has sample mean and standard deviation of 152 and 107, respectively. ARENA input analyzer divided the data into 40 intervals and conducted statistical tests such as Chi-Square and found the best-fit distribution coherent with Erlang. The mathematical expression of the type of distribution was calculated and illustrated below.

$$10 + ERLA(70.8, 2) \tag{4.1}$$

4.3.2.2 Interarrival times

Interarrival time is the time difference between two incoming call dates in a row. We took the unit of interarrival times seconds, as well. Another additional assumption at this point made is eliminating interarrival times that are more than 1000 seconds because the behavior of the raw data is not aligned with that behavior. Afterward, the interarrival times were concatenated in Matlab according to the time interval that they came. By conducting this analysis, we were able to find the interarrival times for each period in 12 hours.

$$InterarrivalTime = CallDate(i+1) - CallDate(i) \tag{4.2}$$

The aim was to find out the number of mean incoming calls for each time period in order to ultimately group and categorize those data according to their volumes. Five categories for the data were created. Those include “very low”, “low”, “middle”, “high”, and “very high”. This grouping was made to make the input data to be analyzed more homogeneous. The illustration of mean incoming call numbers (MICN) with their corresponding categories is below. These groups were determined by analyzing the heatmap given in Figure 4.3. The figures 4.7, 4.8, 4.9, 4.10, 4.11 show the best-fit distributions of these interarrival times.

- 08:30-09:30: “Low”,
- 09:30-10:30: “High”,
- 10:30-15:30: “Very High”,
- 15:30-16:30: “High”,

- 16:30-17:30: “Middle”,
- 17:30-18:30: “Low”,
- 18:30-20:30: “Very Low”.

Figure 4.7 represents the best-fit distribution of the interarrival time category named as very low. As illustrated above, it has a sample mean and standard deviation of 164 and 170 seconds, respectively. ARENA input analyzer divided the input data into 30 intervals and conducted some statistical tests such as Chi-Square and Kolmogorov-Smirnov in order to come up with the type of distribution that the data best fits. Resulted distribution type came out to be Gamma distribution. The full expression is below.

$$-0.001 + GAMM(185, 0.886) \tag{4.3}$$

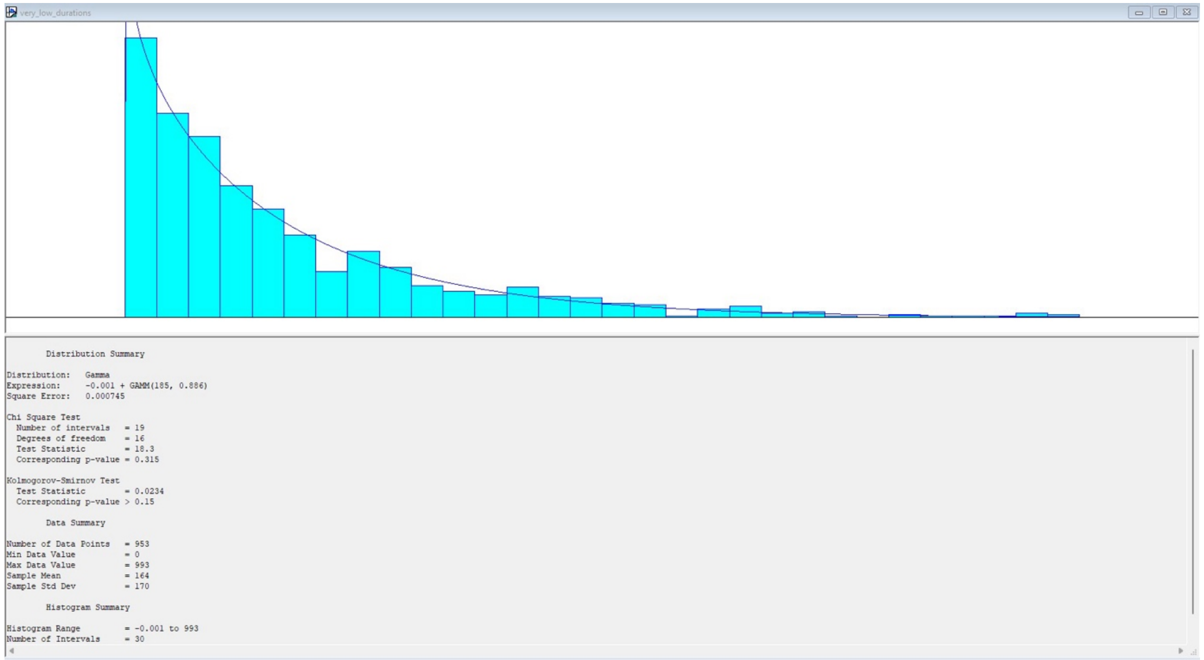

Figure 4.7 : Best-fit distribution of the category "Very Low".

Figures 4.8 and 4.9 represent the best-fit distribution of the interarrival time categories named as low and middle. As illustrated above, they have a sample mean and standard deviation of 94.2 and 102 seconds, and 63.7 and 67.6, respectively.

ARENA input analyzer divided the input data into 40 and 38 intervals for both of these categories in a row and conducted some statistical tests such as Chi-Square and Kolmogorov-Smirnov in order to come up with the type of distribution that the data best fits. Resulted distribution type came out to be exponential distribution. The full expressions are below.

$$-0.001 + EXPO(94.2) \tag{4.4}$$

$$-0.001 + EXPO(63.7) \tag{4.5}$$

Figures 4.10 and 4.11 represent the best-fit distribution of the interarrival time categories named as high and very high. As illustrated above, they have a sample mean and standard deviation of 51.3 and 52 seconds, and 40.4 and 42.1, respectively. ARENA input analyzer divided the input data into 40 and 40 intervals for both of these categories in a row and conducted some statistical tests such as Chi-Square and Kolmogorov-Smirnov in order to come up with the type of distribution that the data best fits. Resulted distribution type came out to be exponential distribution. The full expressions are below.

$$-0.001 + EXPO(51.3) \tag{4.6}$$

$$-0.001 + EXPO(40.4) \tag{4.7}$$

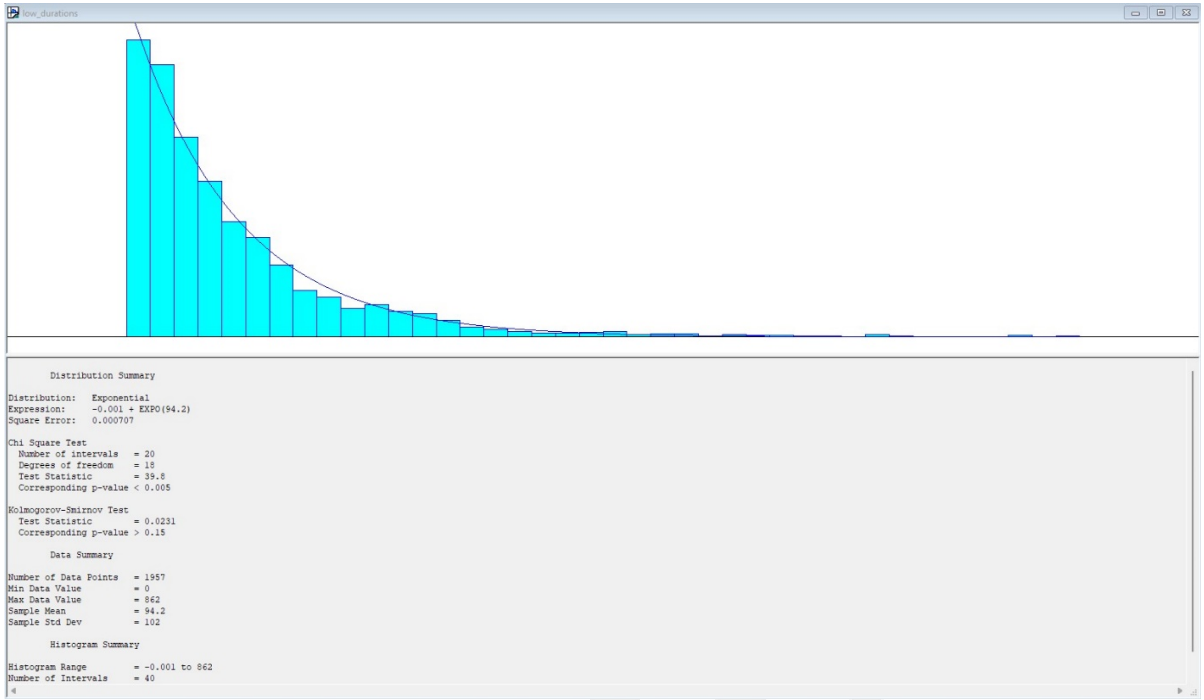

Figure 4.8 : Best-fit distribution of the category “Low”.

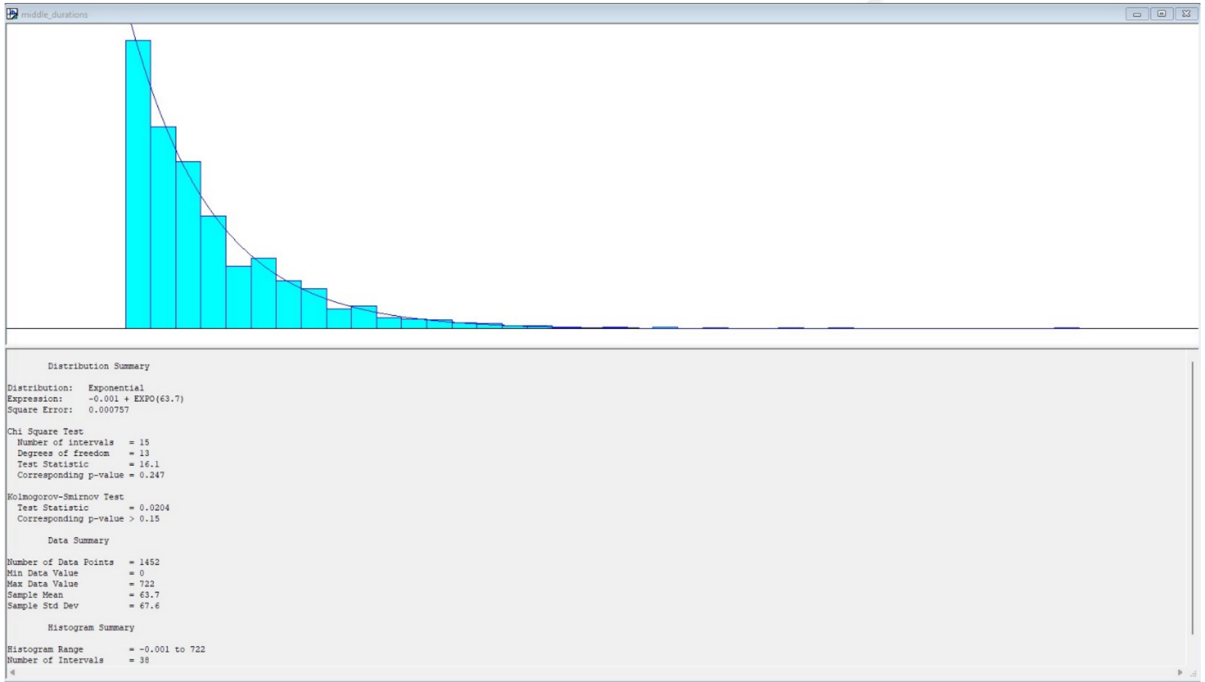

Figure 4.9 : Best-fit distribution of the category “Middle”.

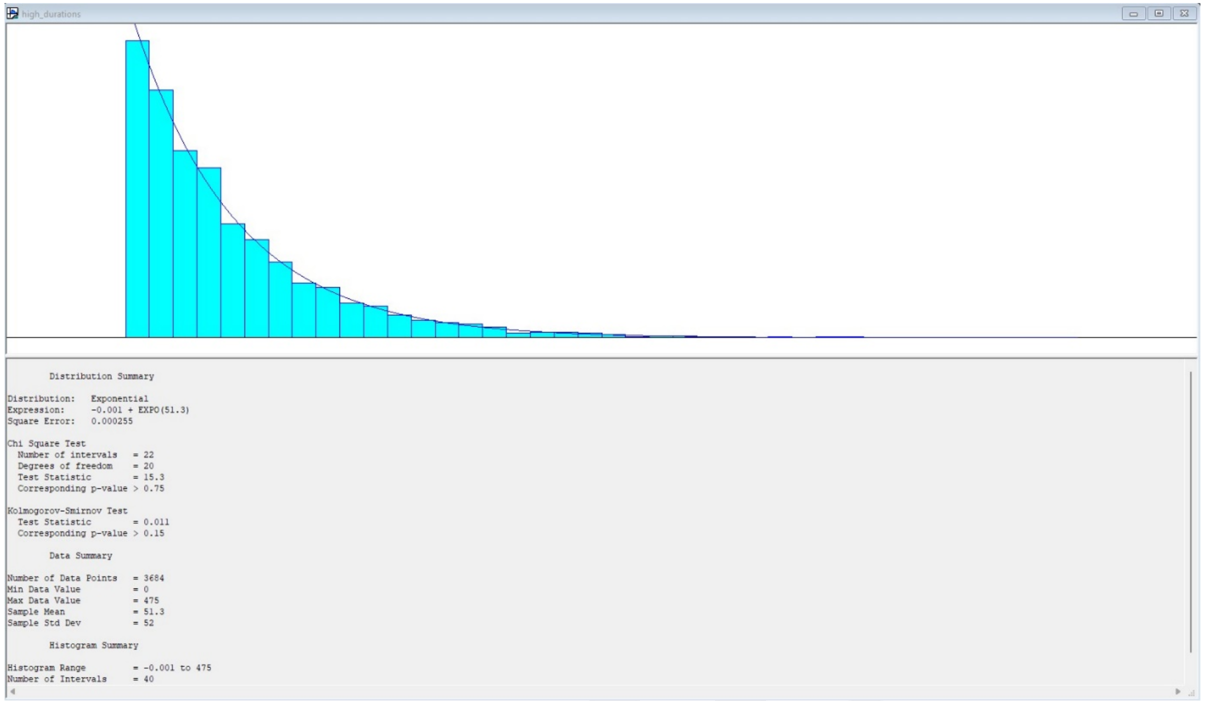

Figure 4.10 : Best-fit distribution of the category “High”.

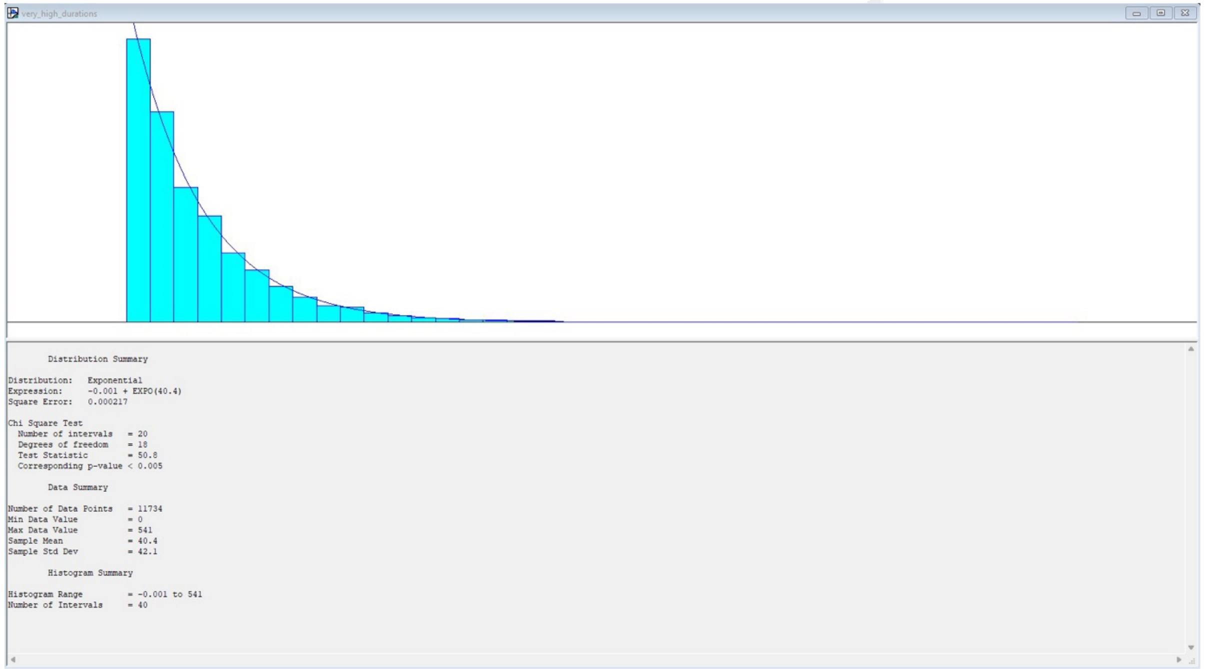

Figure 4.11 : Best-fit distribution of the category “Very High”.

4.3.2.3 Waiting time for missed calls

In figure 4.5, the frequency of lost calls is shown. Lost calls were also put into the ARENA input analyzer tool. Figure 4.12 illustrates the best-fit distribution of the related data in a histogram.

As illustrated in Figure 4.12, it has a sample mean and standard deviation of 68.8 and 58.8 seconds, respectively. ARENA input analyzer divided the input data into 28 intervals and conducted some statistical tests such as Chi-Square and Kolmogorov-Smirnov in order to come up with the type of distribution that the data best fits. Resulted distribution type came out to be exponential distribution. The full expression is below. As mentioned above early in this section, the ones that are less than 10 seconds were eliminated because these calls were defined as immediately shut down, therefore they can not be considered missed calls.

$$-0.001 + EXPO(58.8) \tag{4.8}$$

4.3.2.4 Assignment time to the queue

Assignment to the queue happens immediately after the call arrives. By just observing the raw data shallowly, it can be said that the assignment to the queue happens almost always in a minute. However, for the accuracy and credibility of the simulation application, it needed to be taken into account. This time, we did not have to separate the raw data day by day and hour by hour. The reason behind this is that this is not the place to have any inhomogeneous. So, we only needed the type of distribution of this specific feature.

As seen in Figure 4.13 assignment to the queue data was found to be normally distributed with a mean of 37, and a standard deviation of 7.01.

4.3.2.5 Connection waiting time from queue to operator

In Figure 4.4, the main aim was to observe and determine how the data behaves after extracting the data that is less than 10 seconds. The reason behind this was to find out how many customers made the decision to wait and how many of them dropped the call, according to different time intervals.

On the other hand, differentiating the data by cutting it from the 10-second point and taking both of those separately in the simulation process would not yield accurate results. In addition to this, it is in the nature of the data as well. So, the data also shows the queues, otherwise, there would not be any waiting times more than e few seconds.

At peak times of the day in this cargo call center, it was observed that neither the employees enough nor the customers were satisfied. Bottlenecks occurring in peak times needed to be simulated in investigated in order to come up with a roadmap that solves the problem. This will also be shown in subsection 4.5 where scenario analysis will be explained in detail.

In Figure 4.14, connection waiting times from queue to operator are given. All 20117 data regarding this process were put into the ARENA input analyzer tool. The results showed that it statistically aligns with lognormal distribution with a sample mean of 15.1 and a standard deviation of 34.9.

$$-0.001 + LOGN(10.5, 16.3) \tag{4.9}$$

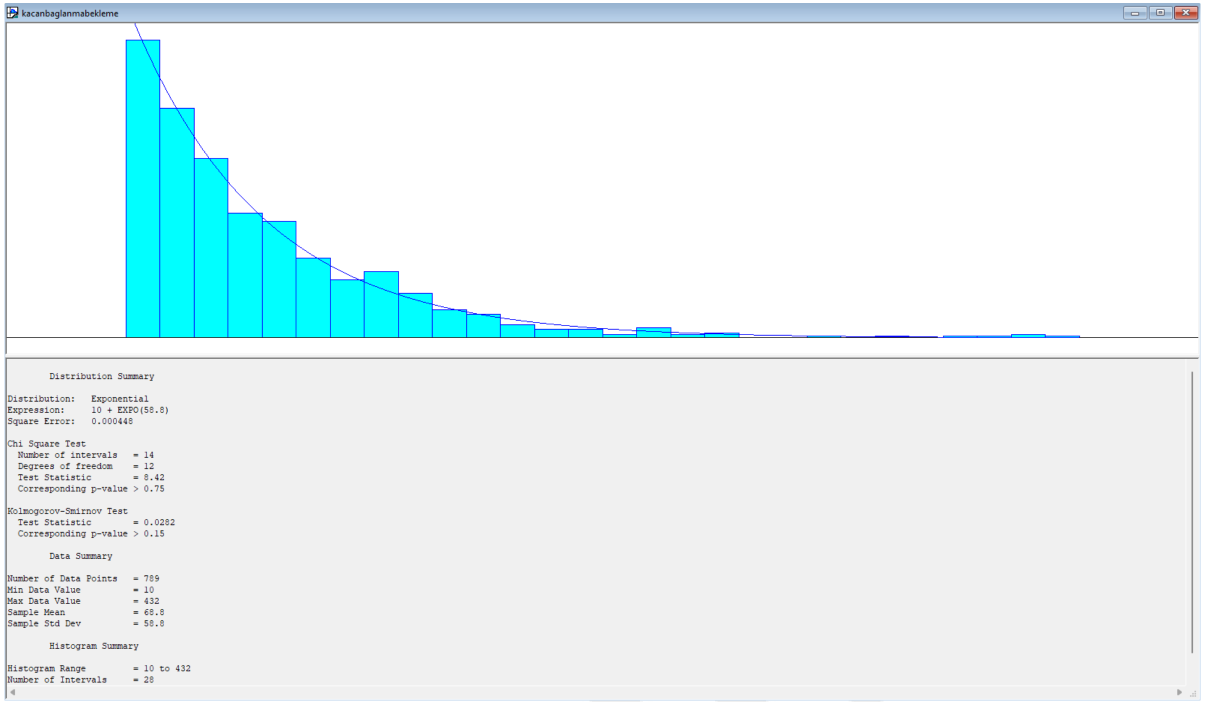

Figure 4.12 : Best-fit distribution of waiting times of missed calls.

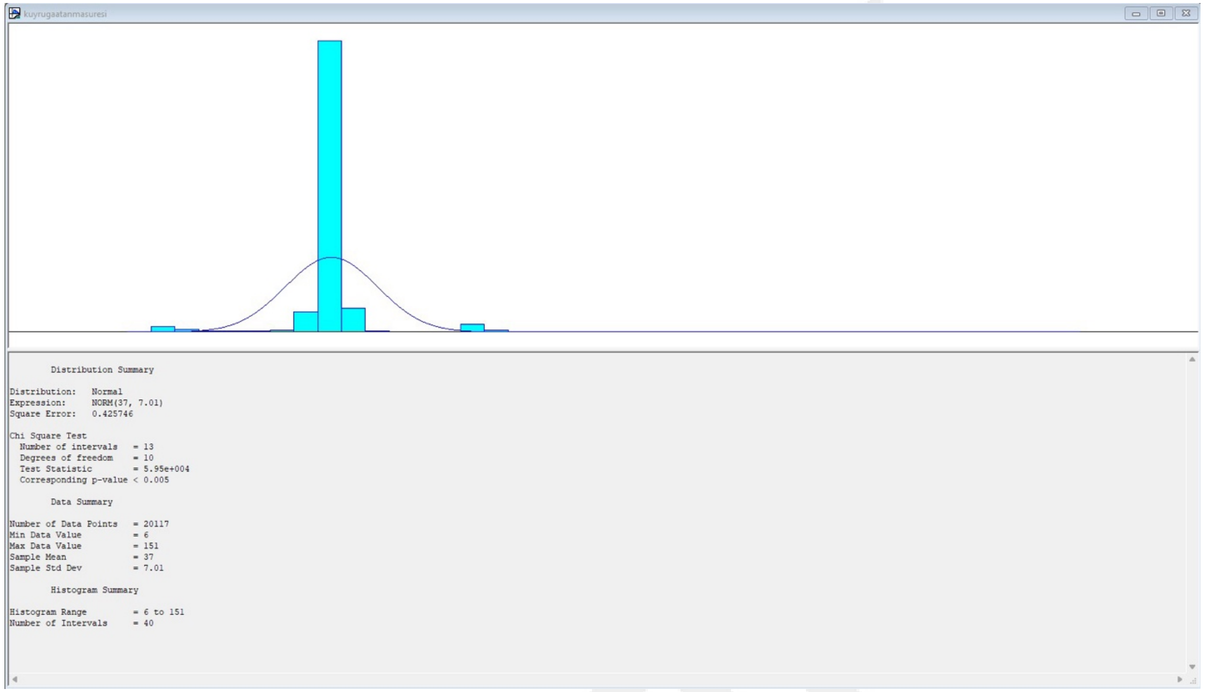

Figure 4.13 : Best-fit distribution of assignment time to the queue.

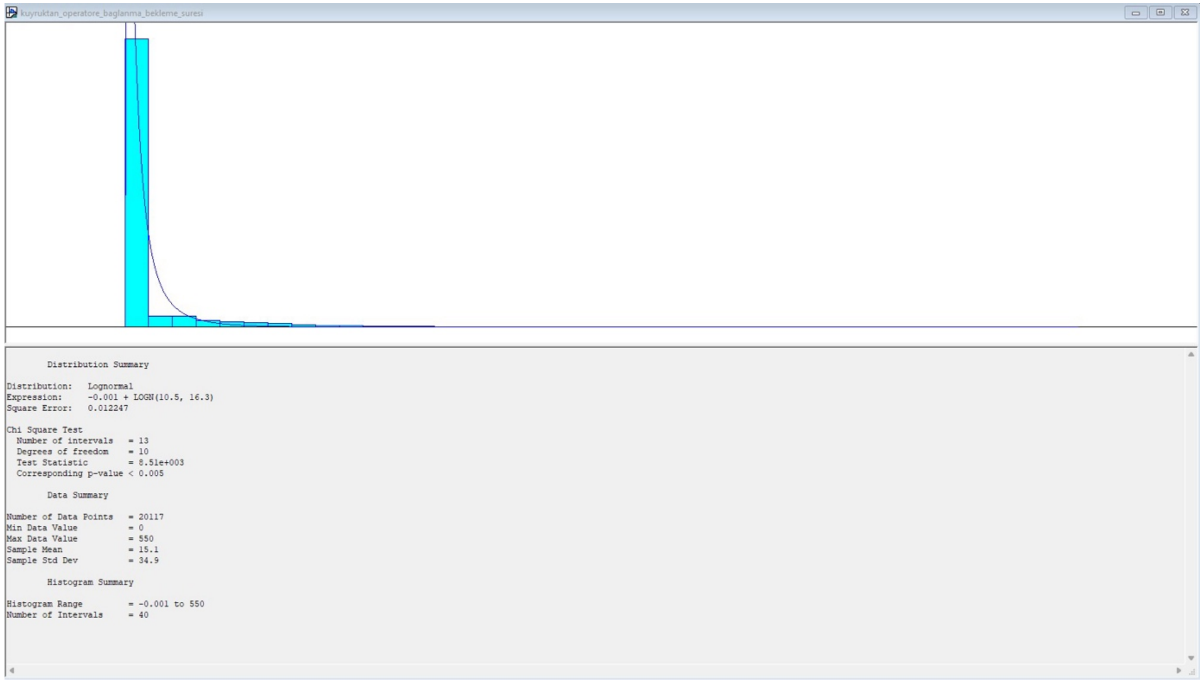

Figure 4.14 : Connection waiting time from queue to operator.

4.4 Process Flow Chart

After conducting data analysis and determining best-fit distributions, the formal model needs to be created according to Figure 4.1 where the framework of this study was determined. The reference model illustrates the main processes and decision points. With the help of a reference model process flow chart is drawn in order to better visualize the system and create the formal model, subsequently.

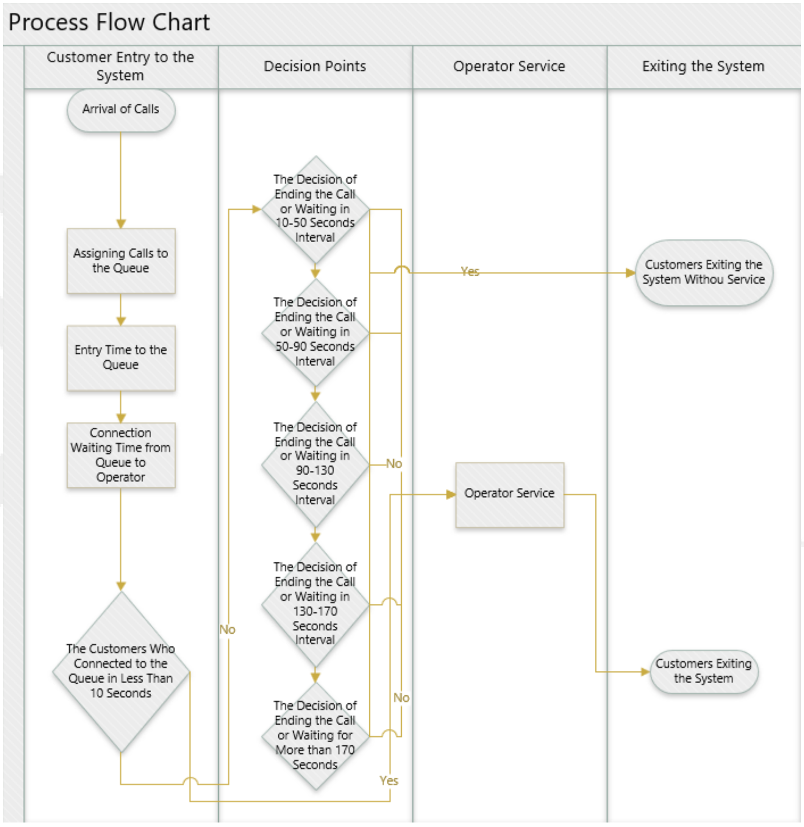

Figure 4.15 : Process flow chart.

Apart from the ARENA model, in Figure 4.15, only the main processes and decision points are shown because some adjustments need to be made in order to set the model correctly and run the simulation program successfully. Additionally, in the ARENA simulation program, some features help build the model and obtain the results correctly. For instance, with the help of an assign module, we can keep track of simulation time and define it as an attribute. It does not only contribute to keeping track of time but also makes it easier to measure the time difference between the processes that were chosen.

Additionally, the record module is generally used to count the number of entities that go by. Saving these is an important factor for both the output report and the scenario analysis.

4.5 Creating Arena Model

ARENA model was built by considering both the reference model and process flow chart. When the formal model is created, it is also important to validate this model with the actual service system. If there exist any misleading points that need to be extracted, the model must be re-created by going backward towards even the data collection step. Validation is the step where the analyst checks the model against the actual system and compares those in order to see whether the formal model accurately represents the overall system or not.

By using the basic process panel in ARENA, the formal model is created. Figure 4.16 represents the formal model. The formal model occupies too much place. Therefore, firstly the whole model will be given. Afterwards, it will be given in sections.

The arrival of calls is created with the create module in the ARENA program. The key point in this section was to successfully separate and unite the raw data. Raw data had the incoming call dates. Moreover, it was reorganized so as to enter every type of best-fit distribution in their corresponding time intervals. For instance, in the first part which is illustrated as the arrival of calls, the incoming calls in the time interval of 08:30-09:30 for 27 days come first. Then, the other incoming calls are taken to the system, subsequently.

Interarrival times were determined for every period of a working day in this cargo call center. The incoming calls were divided into 5 categories that align with the incoming volumes of calls. As explained before, assigning calls to the queue is another process that needs to be taken into account. Going forward, the third object is an assigned module that was placed there in order to keep track of simulation time by defining an attribute and assigning the TNOW function to it.

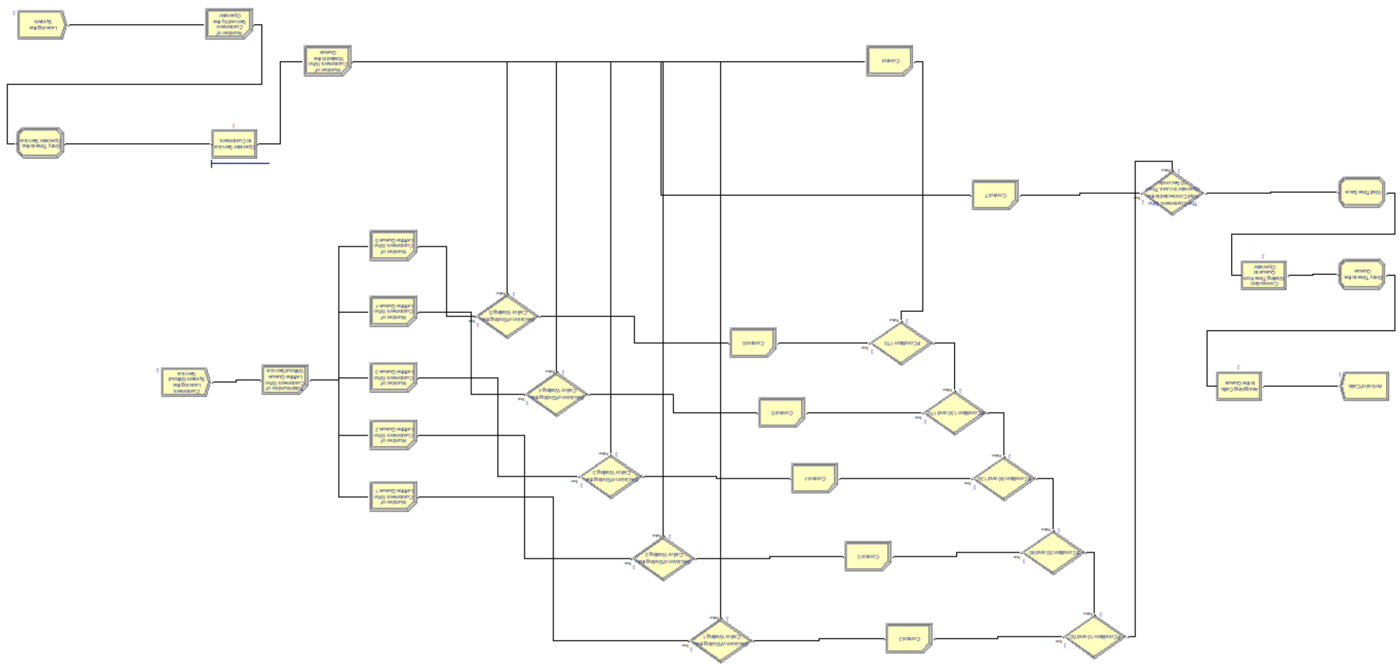

Figure 4.16 : ARENA model.

All the processes apart from interarrival time have one, unique type of statistical distribution. However, because the interarrival time data showed inhomogeneity, dividing it into five different categories according to incoming call volumes was necessary to have accurate results at the end. On the contrary, connection waiting time from the queue to the operator has its unique best-fit distribution.

It is seen from Figure 4.17 that there is an assigned module used again with the name given as wait time save. This is the part where the process flow chart does not exactly match the formal model because some arrangements need to be set in order to run the simulation program.

In Figures 4.4 and 4.5, it was explained that connection waiting time from the queue to the operator is critical because if it is a peak time of the day, there might be some unintended delays due to a lack of employees and high demand, which also would create long queues. Before overcoming this issue, it first needs to be represented correctly.

For that reason, it is beneficial to point out that with the help of the assigned module, named wait time save, is able to keep the value-added times of each entity which behaves statistically as connection waiting time from queue to operator process. Defining this assignment helps also define the decision points separately, and accurately.

The customers who connected to the queue in less than 10 seconds are already going towards the operator service process. However, the ones that waited more than ten seconds show different characteristics. Long waiting times in queues do not align with any company's interests because they cause customer dissatisfaction which again trying to prevent it is one of the main aims here.

After analyzing the raw data, it was seen that there are some customers who decide to wait no matter how much it takes, and there are some customers who decide whether to wait or leave by taking the time they waited into consideration. Finally, there exists also another type of customer that is mostly impatient.

The customers who wait by taking the time to wait into consideration also show different characteristics among them. For instance, approximately 50% of customers

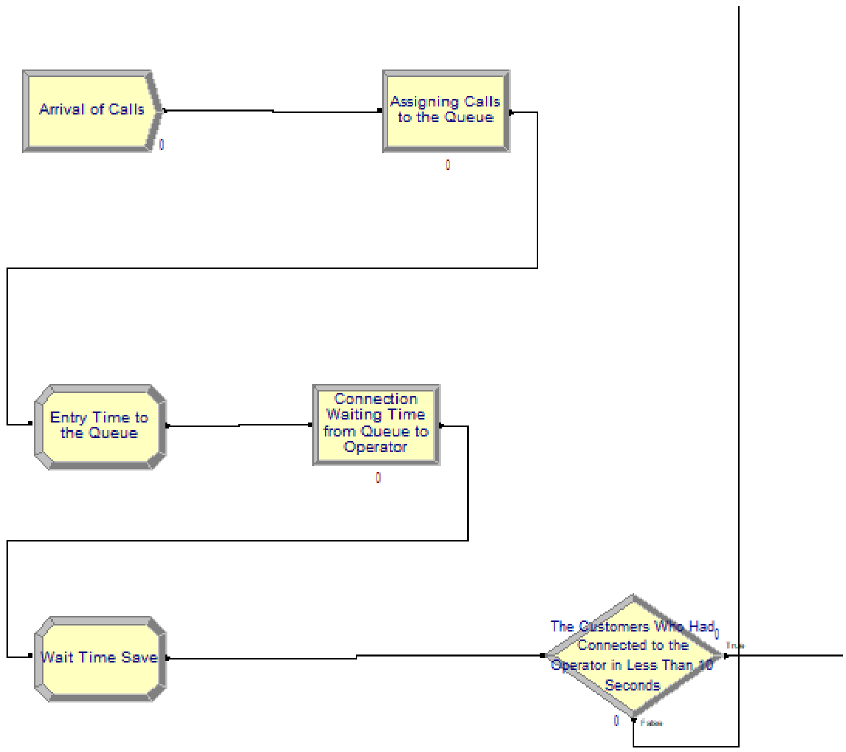

Figure 4.17 : First section of ARENA model.

who choose to leave the queue without taking any service, only wait for 10 to 50 seconds according to Figure 4.5.

That is the reason why a two-by-chance decision module in ARENA was needed. If the customers cannot be connected to the operator in less than ten seconds, the ones in the queue will also be separated by looking at the statistical properties of those. Imagining a time interval is beneficial for better understanding the process. Dividing the lost or missed calls by the total number of calls that are again in that interval, being able to determine a probability becomes easy for customers about which path they will take in terms of deciding whether to wait or not.

In Figure 4.18, these decision points are shown. There are five different decision points for five different wait time intervals. The calculation steps that were explained before, are used here in order to represent the system exactly, so as to validate the model created. Table 4.2 shows the calculated probabilities for ending the call decision.

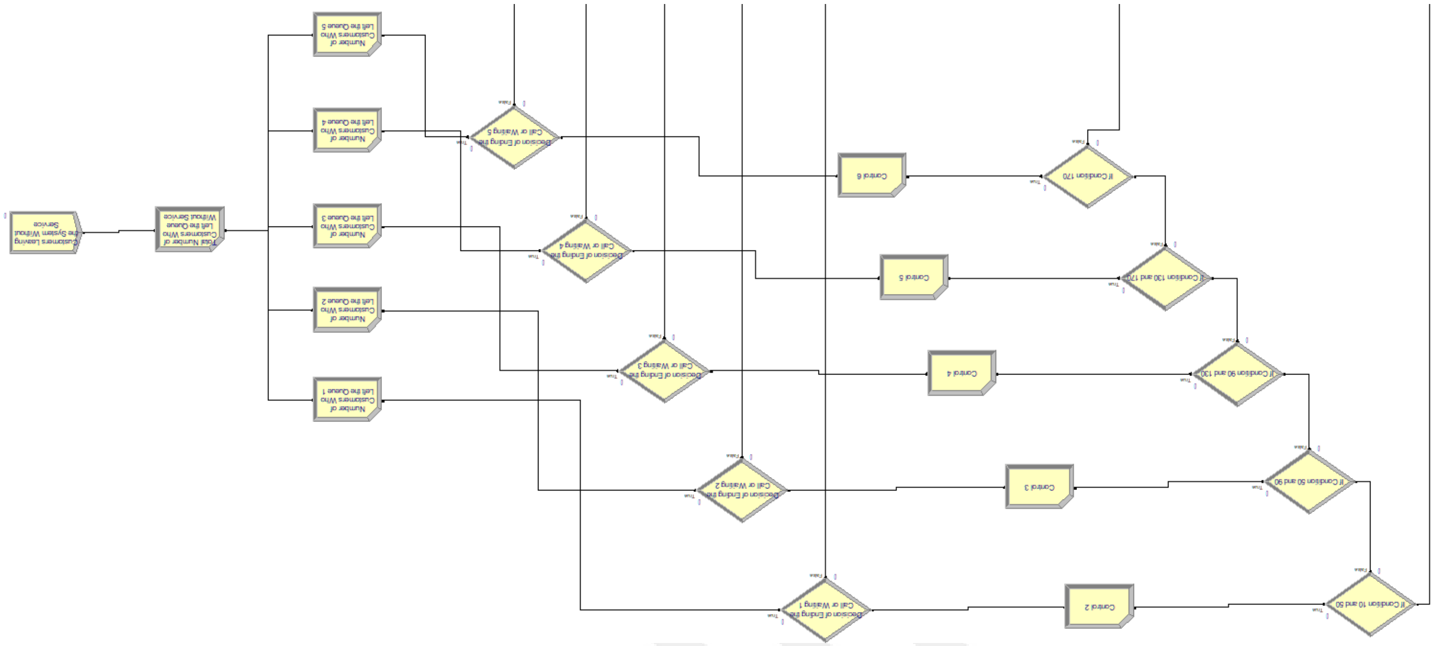

Figure 4.18 : Second section of ARENA model.

Table 4.2 : Decision probability of leaving.

Ending/Waiting in	Type of Decision Defined	Chance of Ending the Call
10-50 seconds	2-way by Chance	21.41%
50-90 seconds	2-way by Chance	21.09%
90-130 seconds	2-way by Chance	24.44%
130-170 seconds	2-way by Chance	23.04%
>170 seconds	2-way by Chance	20.45%

The other way in these decision points goes towards the operator. Those who stayed in the queue then being directed to the operator to be served. Record modules in Figure 4.18 again to keep the statistics for the output report. In the last section of the ARENA model, the total number of customers who waited in the queue is recorded to verify the output results. And then, operator service comes where the queue actually occurs. In the cargo call center, the employees were working with different shifts that were aiming to maximize the number of employees working in peak hours and minimize otherwise. For each time interval, there are different numbers of employees working. Therefore, similarly to the interarrival times, operators needed to be scheduled, too. Finally, the simulation program ends with recording also the statistics after the operator service and leaving the system. Figure 4.19 shows the third and last section of the simulation program in detail.

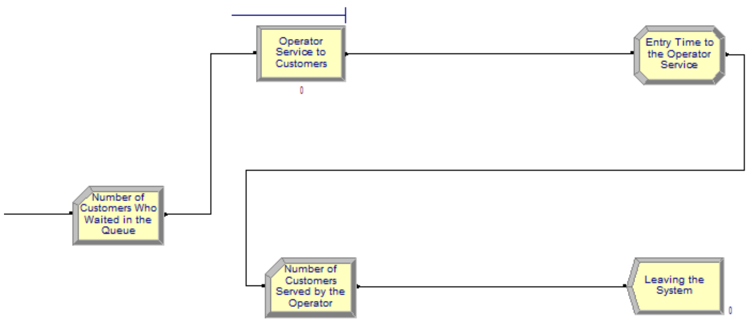

Figure 4.19 : Third section of ARENA model.

Setting up simulation operation parameters

All necessary parameters that are used in the simulation operation are in units of seconds. There is no warm-up period specified. The call center is operational for

Table 4.3 : Utilization rates for initial replications.

Number of Replications	Instantaneous Utilization
1	0.0626
2	0.2427
3	0.2481
4	0.3251

12 hours 6 days a week. The calculated replication length for each day corresponds to 43200 seconds.

Another important calculation that needs to be done is to find out how many replications the data needs. In theory, the more the replication number is, the less the corresponding error will be. Therefore, high numbers of replications are generally preferred when conducting simulation analysis. However, it is also important to point out the minimum number of replications needed. To calculate the minimum number of replications needed, the analyst needs to determine the confidence interval, sample variance of utilization rates for at least the first few replications, error criterion(ε), and t table values. The related data is shown in Table 4.2 and Table 4.3.

ε was determined as 0.04. Because the desired confidence interval is 95%, $\alpha/2$ yields to 0.025. These calculations also aim to acquire reliable results. Otherwise, with an insufficient number of replications, half-widths would not be calculated. Generally, in simulation applications, independent replications are used, each run using a different random number stream and independently chosen initial conditions.

$$R \geq \left(\frac{t_{(\frac{\alpha}{2}, R-1)} \cdot S_{(0)}}{\varepsilon} \right)^2 \tag{4.10}$$

As seen on Table 4.4, in replication 33, $R \geq R_0$ condition is met. Therefore, the minimum number of replications needed for this simulation study is calculated as 33. Any number of replications higher than this number is also sufficient and enough. In the simulation application, the number of replications was set to 1000 for all scenarios.

4.6 Output & Scenario Analysis

The initial number of employees was known. The simulation ran for 1000 replications. As explained before, the employees work in different shifts. For that reason, every

Table 4.4 : Calculation of sufficient number of replications.

R	$R \geq \left(\frac{t_{\left(\frac{\alpha}{2}, R-1\right)} \cdot S_{(0)}}{\varepsilon} \right)^2$	R_0
1	-	-
2	12.70	1248.70
3	4.30	143.18
4	3.18	78.33
5	2.77	59.62
6	2.57	51.10
7	2.44	46.30
8	2.36	43.24
9	2.30	41.12
10	2.26	39.57
11	2.22	38.39
12	2.20	37.46
13	2.17	36.71
14	2.16	36.09
15	2.14	35.57
11	2.22	38.39
12	2.20	37.46
13	2.17	36.71
14	2.16	36.09
15	2.14	35.57
16	2.13	35.13
17	2.11	34.75
18	2.10	34.42
19	2.10	34.13
20	2.09	33.88
21	2.08	33.65
22	2.07	33.44
23	2.07	33.26
24	2.06	33.09
25	2.06	32.94
26	2.05	32.80
27	2.05	32.67
28	2.05	32.56
29	2.04	32.45
30	2.04	32.35
31	2.04	32.25
32	2.03	32.17
33	2.03	32.09
34	2.03	32.01

employee that works in different time intervals is determined first, representing the current system as is.

On the other hand, three additional scenarios were added to the scenario analysis in order to better achieve the desired level of satisfaction through mainly decreasing the utilization rate, in other words, relaxing the bottleneck.

All KPIs and important data that were measured are listed below.

- Utilization
- Value-Added Time per Entity (Process)
- Wait Time per Entity (Process)
- Total Time per Entity (Process)
- Waiting Time (Queue)
- Total Time in the System (In Seconds)
- Idle Position
- Number of Customers Who Left the Queue Without Service
- Number of Customers Who Waited in the Queue
- Missed Call Ratio

All scenarios except Scenario-1 are hypothetical and aim for effectively and efficiently achieving better performance in terms of KPIs. For example, increasing the number of employees at peak hours of the day should decrease the utilization rate and relax the bottleneck hypothetically.

The employees that are working intentionally differentiated according to the time interval they work in. The reason behind this is that the day was divided into several categories based on the incoming call volumes. In order to have a better understanding and analysis, those categories are strictly followed in the scenario analysis.

As expected, the utilization rate decreased when the number of employees who work at peak hours increased in Table 4.5. Moreover, when value-added times per entity

Table 4.5 : Utilization rates of scenarios & VA times per entity.

Scenarios	Utilization	VA Time Per Entity(Process)(In Seconds)		
		Assigning Calls to the Queue	Connection Waiting Time from Queue to Operator	Operator Service to Customers
Scenario-1	38.21%	36.99	10.48	151.46
Scenario-2	33.33%	37.00	10.46	151.45
Scenario-3	31.17%	36.99	10.46	151.67
Scenario-4	27.08%	36.98	10.46	151.35

are analyzed there are no significant changes. This is because these processes follow a steady state, meaning, the expected utilization rate decreased because the queue was relaxed by the increased number of employees at peak hours, however, this scenario would not change the time spent in operator service because it depends on the data that represents the raw interaction customer and employee have.

Significant changes occurred, as expected, in waiting times in both the queue and the system. Total time per entity in processes decreased from 753.85 seconds to 486.12 seconds from Scenario-1 to Scenario-4, which is a crucial improvement in terms of customer satisfaction. Similarly, waiting time in queue decreased from 561.77 seconds to 295.28 seconds. As a result, the changes made in scenarios decreased the total time spent in the system from 1315.62 seconds to 781.40 seconds.

So far, the objectives set before the simulation application have been met. KPIs are accurately measured to improve the overall performance. Secondly, the staff is scheduled effectively and efficiently. Finally, customer satisfaction is reached when the waiting times in Table 4.7 are taken into account. However, other parameters are also important to mention.

Table 4.6 : Number of employees who worked in different time intervals.

Scenario	8:30 - 9:00	9:00 - 10:00	10:00 - 11:00	11:00 - 17:00	17:00 - 18:00	18:00 - 19:00	19:00 - 20:00	20:00 - 20:30
Scenario-1	2	4	7	9	7	5	3	2
Scenario-2	3	5	9	13	9	4	3	2
Scenario-3	3	6	10	14	10	5	3	2
Scenario-4	3	9	13	16	13	7	3	3

Table 4.7 : Waiting & total times (in seconds).

Scenarios	Wait Time Per Entity(Process)	Waiting Time	Total Time Per Entity(Process)	Total Time
Scenario-1	554.92	561.77	753.85	1315.62
Scenario-2	462.35	470.01	661.26	1131.27
Scenario-3	382.64	389.70	581.76	971.46
Scenario-4	287.33	295.28	486.12	781.40

Table 4.8 shows the idle position of employees and a number of statistical data that were collected for the output analysis. As seen from Table 7, the idle position of employees increased. It is a reasonable change because the bottleneck and employee utilization were relaxed.

Table 4.8 : Other statistical outputs.

Scenarios	Idle Position	Number of Customers Who Left the Queue Without Service	Number of Customers Waited in the Queue	Missed Call Ratio
Scenario-1	4.78	135.81	749.86	18.11%
Scenario-2	7.17	134.49	749.46	17.94%
Scenario-3	7.98	132.86	752.09	17.67%
Scenario-4	9.87	133.34	749.77	17.78%

As shown in Table 4.8, the idle position of employees increased from Scenario-1 to Scenario-4. Missed call ratio did not show a significant change due to the nature of the behavior of raw data. The number of customers who left the queue without service and who on the other hand stayed, showed similar bahavior in the same time intervals illustrated in Figures 4.4 and 4.5.

5. CONCLUSIONS

Call centers play a crucial role in the service sector, especially in B2B and B2C operations. Direct customer-operator interaction occurs on a frequent basis. Aiming for an increased level of efficiency and customer satisfaction are the critical objectives to be reached.

Almost all real-life problems regarding call center operations have their own specific nature. When real-life instances are considered, most of the data shows stochastic characteristics. Deterministic approaches to solving problems almost always yield wrong outcomes because these approaches or methods are developed with a high number of assumptions that were made to limit the unforeseen events in nature so as to come up with a result. However, limiting the events occurring in everyday life would also limit and decrease the level of accuracy of that result.

Many applications have been conducted so far regarding the call center staff scheduling problems in which various methods were used such as simulation, optimization, forecasting, and various combinations of these. In this study, the call center staff scheduling problem is solved with simulation, namely, the discrete event simulation technique.

The problem arises from a number of employees that fall short at peak times in working days of the call center. Long waiting times in queue occurred, hence, these raised the problem of customer dissatisfaction. Bottlenecks occurring at peak hours need to be taken into account seriously in order to improve the overall system dynamics. Adapting the existing resources accordingly would improve the system in which the simulation as a tool intervenes.

Simulation is a powerful tool when its capabilities are considered. Conceptually creating a reference model, developing it, and finally creating a formal model is one side. The other side is that simulation as a tool allows the analyst to see the system as a whole comprehensively, and manage and monitor it.

In the problem, the call center had a single channel operating. Several processes were being conducted in the customer interaction. The main problems were managerially defined listed below.

- Not being able to measure KPI's accurately.
- Not being able to schedule staff effectively and efficiently.
- Bottleneck occuring at peak hours.

In order to better understand the current system operation, data collection, model conceptualizing, and reference model creation have been studied at the same time. The raw data acquired from the firm had 20117 lines of incoming calls in March 2022. There were several different columns that stand for different recordings of incoming calls such as the exact time that an incoming call was assigned to the queue.

By conducting some mathematical calculations in Matlab and drawing the data that is ready to use in the simulation program ARENA as an input, best-fit distributions of all necessary data were found. Subsequently, the formal model was created and validated by also comparing the actual system with the model created.

In simulation analysis, the number of replications plays a crucial role in terms of reliable outcomes. Misleading results are more likely to occur if the minimum necessary number of replications is not reached. In the output report in the ARENA simulation program, for every KPI, the program tries to calculate the half-widths. Another way of finding out if the necessary minimum number of replications is reached or not is to look at the half-widths. If those were not calculated by the program itself, that also means that the number of replications is not enough.

Therefore, the necessary minimum number of replications was calculated and yielded to the number 33, meaning, any number of replications that is higher than 33 will yield accurate and reliable results.

However, in any case, the number of replications used in this simulation was set to 1000. The firm was operating for 12 hours for six days a week. In the formal simulation model, every entity and process was measured in seconds in terms of time units. Then, output and scenario analysis were conducted. The time that every entity, customer, spends in both the queue and the system was decreased in order to achieve customer satisfaction. Additionally, the bottleneck was relaxed due to the change made in scenario analysis regarding the employees working in each time period, namely shift. Successfully decreasing the utilization rate which also increased the idle position of employees helped achieve one of the objectives.

Moreover, another important thing to mention is that the KPIs were successfully measured and calculated for verification. Going forward, the bottleneck was also relaxed due to the changes made. So, all of the managerial objectives were achieved in a concrete, solid simulation analysis and application.

In the literature review, numerous studies that optimization as a tool was used along with simulation and forecasting. For future studies, optimization can be used for even more improved and accurate results.

In optimization, there are three types of solutions which are as follows:

- Feasible Solution
- Optimum Solution
- Near-Optimum Solution

For real-life instances, it is often difficult to obtain optimum results because the volume of the data is huge compared to experimental studies. Additionally, it is not always possible to satisfy all of the objectives, which otherwise would yield to feasible solution.

In that case, trying to achieve a near-optimum solution is mostly desired when both the time limitations and volume of raw data are taken into consideration. Companies often prefer more quick, effective, and efficient solutions rather than exactly feasible ones.

Furthermore, exact methods in optimization are not practical when real-life problems are considered. Therefore, approximate methods can be used, namely, metaheuristics, along with simulation.

Due to the NP-Hard nature of the models developed to solve the staff scheduling problem, meteheuristic algorithms are widely used to solve real-life problems. By using these algorithms, it is aimed to obtain near-optimal results in reasonable run time. In future studies, mathematical models specific to the nature of the problem can be developed for the staff scheduling problem in call centers and models developed with different metaheuristic algorithms can be solved.

Metaheuristic algorithms have been developed with inspiration from positive sciences or from nature. The aim of these algorithms is to avoid the local optimum traps and try to find out global optimum solution in the shortest possible time. The solution type these algorithms try to achieve is the near-optimum solution. Commonly known metaheuristic algorithms are as follows:

- Tabu Search Algorithm,
- Genetic Algorithm,
- Ant Colony Algorithm,
- Particle Swarm Algorithm,
- Artificial Bee colony Algorithm, etc.

When optimization techniques can not find the solution to the problem of interest, simulation as a tool generally being employed. Simulation is employed when the system is complex and hard to define in mathematical models and equations. Simulation is the imitation of the operation of a real-life process or system over time and can be used as an analysis tool for predicting the effect of changes to existing systems.

Simulation is mainly neither a solution nor an answer, but an iterative experiment on the system of interest. Therefore, no optimum results are achieved by employing simulation as a tool. It can be described as more or less an input-output model.

If the system of interest is too complex, and understanding the effects of changes made is hard, simulation could be an appropriate tool to use. If a problem can be analytically solved, simulation is not needed and is also not an appropriate tool to use. The main areas of simulation are being applied are as follows:

- Manufacturing Applications,
- Healthcare Applications,
- Logistics, Transportation, and Distribution Applications,
- Business Process Simulation,
- Service Systems.

REFERENCES

[1] **Thompson, G.M.** (1996). A simulated-annealing heuristic for shift scheduling using non-continuously available employees, *Computers & Operations Research*, *23*(3), 275–288.

[2] **Easton, F.F. and Mansour, N.** (1999). A distributed genetic algorithm for deterministic and stochastic labor scheduling problems, *European Journal of Operational Research*, *118*(3), 505–523.

[3] **Cai, X. and Li, K.** (2000). A genetic algorithm for scheduling staff of mixed skills under multi-criteria, *European journal of operational research*, *125*(2), 359–369.

[4] **Koole, G. and Van Der Sluis, E.** (2003). Optimal shift scheduling with a global service level constraint, *IIE Transactions*, *35*(11), 1049–1055.

[5] **Bard, J.F.** (2004). Staff scheduling in high volume service facilities with downgrading, *Iie Transactions*, *36*(10), 985–997.

[6] **Robbins, T.R. and Harrison, T.P.** (2010). A stochastic programming model for scheduling call centers with global service level agreements, *European Journal of Operational Research*, *207*(3), 1608–1619.

[7] **Ingolfsson, A.**, **Campello, F.**, **Wu, X. and Cabral, E.** (2010). Combining integer programming and the randomization method to schedule employees, *European Journal of Operational Research*, *202*(1), 153–163.

[8] **Brucker, P.**, **Qu, R. and Burke, E.** (2011). Personnel scheduling: Models and complexity, *European Journal of Operational Research*, *210*(3), 467–473.

[9] **Van den Bergh, J.**, **Beliën, J.**, **De Bruecker, P.**, **Demeulemeester, E. and De Boeck, L.** (2013). Personnel scheduling: A literature review, *European journal of operational research*, *226*(3), 367–385.

[10] **Gabrel, V.**, **Murat, C. and Thiele, A.** (2014). Recent advances in robust optimization: An overview, *European journal of operational research*, *235*(3), 471–483.

[11] **Lin, S.W. and Ying, K.C.** (2014). Minimizing shifts for personnel task scheduling problems: A three-phase algorithm, *European Journal of Operational Research*, *237*(1), 323–334.

[12] **Excoffier, M., Gicquel, C. and Jouini, O.** (2016). A joint chance-constrained programming approach for call center workforce scheduling under uncertain call arrival forecasts, *Computers & Industrial Engineering*, *96*, 16–30.

[13] **Walter, M. and Zimmermann, J.** (2016). Minimizing average project team size given multi-skilled workers with heterogeneous skill levels, *Computers & Operations Research*, *70*, 163–179.

[14] **Taskiran, G.K. and Zhang, X.** (2017). Mathematical models and solution approach for cross-training staff scheduling at call centers, *Computers & Operations Research*, *87*, 258–269.

[15] **Mattia, S., Rossi, F., Servilio, M. and Smriglio, S.** (2017). Staffing and scheduling flexible call centers by two-stage robust optimization, *Omega*, *72*, 25–37.

[16] **Bodur, M. and Luedtke, J.R.** (2017). Mixed-integer rounding enhanced benders decomposition for multiclass service-system staffing and scheduling with arrival rate uncertainty, *Management Science*, *63*(7), 2073–2091.

[17] **Vermuyten, H., Rosa, J.N., Marques, I., Belien, J. and Barbosa-Póvoa, A.** (2018). Integrated staff scheduling at a medical emergency service: An optimisation approach, *Expert systems with applications*, *112*, 62–76.

[18] **Xu, Y. and Wang, X.** (2021). An artificial bee colony algorithm for scheduling call centres with weekend-off fairness, *Applied Soft Computing*, *109*, 107542.

[19] **Xu, Y. and Wang, X.** (2022). A hybrid integer programming and artificial bee colony algorithm for staff scheduling in call centers, *Computers & Industrial Engineering*, *171*, 108312.

[20] **Boysen, N., Emde, S. and Schwerdfeger, S.** (2022). Crowdshipping by employees of distribution centers: Optimization approaches for matching supply and demand, *European Journal of Operational Research*, *296*(2), 539–556.

[21] **Mansini, R., Zanella, M. and Zanotti, R.** (2023). Optimizing a complex multi-objective personnel scheduling problem jointly complying with requests from customers and staff, *Omega*, *114*, 102722.

[22] **Ernst, A.T., Jiang, H., Krishnamoorthy, M. and Sier, D.** (2004). Staff scheduling and rostering: A review of applications, methods and models, *European journal of operational research*, *153*(1), 3–27.

[23] **Mehrotra and Fama** (2003). Call center simulation modeling: methods, challenges, and opportunities, *Proceedings of the 2003 Winter Simulation Conference, 2003.*, volume 1, IEEE, pp.135–143.

[24] **Whitt, W.** (2005). Engineering solution of a basic call-center model, *Management Science*, *51*(2), 221–235.

[25] **Feldman, Z., Mandelbaum, A., Massey, W.A. and Whitt, W.** (2008). Staffing of time-varying queues to achieve time-stable performance, *Management Science*, *54*(2), 324–338.

[26] **Doomun, R. and Vunka Jungum, N.** (2008). Business process modelling, simulation and reengineering: call centres, *Business Process Management Journal*, *14*(6), 838–848.

[27] **Ma, J., Kim, N. and Rothrock, L.** (2011). Performance assessment in an interactive call center workforce simulation, *Simulation Modelling Practice and Theory*, *19*(1), 227–238.

[28] **Titler, S.S., Dexter, F. and Epstein, R.H.** (2021). Impact of anesthesia resident staff assignment decisions on nurse anesthetist and anesthesia resident staff scheduling and productivity: tutorial using data from a pediatric hospital, *Perioperative care and operating room management*, *24*, 100182.

[29] **Taylor, J.W.** (2012). Density forecasting of intraday call center arrivals using models based on exponential smoothing, *Management Science*, *58*(3), 534–549.

[30] **Ibrahim, R., Ye, H., L'Ecuyer, P. and Shen, H.** (2016). Modeling and forecasting call center arrivals: A literature survey and a case study, *International Journal of Forecasting*, *32*(3), 865–874.

[31] **Tang, X., Liao, X., Zheng, J. and Yang, X.** (2018). Energy efficient job scheduling with workload prediction on cloud data center, *Cluster Computing*, *21*, 1581–1593.

[32] **Chacón, H., Koppisetti, V., Hardage, D., Choo, K.K.R. and Rad, P.** (2023). Forecasting call center arrivals using temporal memory networks and gradient boosting algorithm, *Expert Systems with Applications*, *224*, 119983.

[33] **Atlason, J., Epelman, M.A. and Henderson, S.G.** (2004). Call center staffing with simulation and cutting plane methods, *Annals of Operations Research*, *127*, 333–358.

[34] **Wallace, R.B. and Whitt, W.** (2005). A staffing algorithm for call centers with skill-based routing, *Manufacturing & Service Operations Management*, *7*(4), 276–294.

[35] **Harrison, J.M. and Zeevi, A.** (2005). A method for staffing large call centers based on stochastic fluid models, *Manufacturing & Service Operations Management*, *7*(1), 20–36.

[36] **Aksin, Z., Armony, M. and Mehrotra, V.** (2007). The modern call center: A multi-disciplinary perspective on operations management research, *Production and operations management*, *16*(6), 665–688.

[37] **Deslauriers, A., L'Ecuyer, P., Pichitlamken, J., Ingolfsson, A. and Avramidis, A.N.** (2007). Markov chain models of a telephone call center with call blending, *Computers & operations research*, *34*(6), 1616–1645.

[38] **Ertogral, K. and Bamuqabel, B.** (2008). Developing staff schedules for a bilingual telecommunication call center with flexible workers, *Computers & Industrial Engineering*, *54*(1), 118–127.

[39] **Cezik, M.T. and L'Ecuyer, P.** (2008). Staffing multiskill call centers via linear programming and simulation, *Management Science*, *54*(2), 310–323.

[40] **Atlason, J., Epelman, M.A. and Henderson, S.G.** (2008). Optimizing call center staffing using simulation and analytic center cutting-plane methods, *Management Science*, *54*(2), 295–309.

[41] **Chevalier, P. and Van den Schrieck, J.C.** (2008). Optimizing the staffing and routing of small-size hierarchical call centers, *Production and Operations Management*, *17*(3), 306–319.

[42] **van Dijk, N.M. and van der Sluis, E.** (2008). Practical optimization by OR and simulation, *Simulation Modelling Practice and Theory*, *16*(8), 1113–1122.

[43] **Castillo, I., Joro, T. and Li, Y.Y.** (2009). Workforce scheduling with multiple objectives, *European Journal of Operational Research*, *196*(1), 162–170.

[44] **Avramidis, A.N., Chan, W., Gendreau, M., L'ecuyer, P. and Pisacane, O.** (2010). Optimizing daily agent scheduling in a multiskill call center, *European Journal of Operational Research*, *200*(3), 822–832.

[45] **Bertsimas, D. and Doan, X.V.** (2010). Robust and data-driven approaches to call centers, *European Journal of Operational Research*, *207*(2), 1072–1085.

[46] **Dietz, D.C.** (2011). Practical scheduling for call center operations, *Omega*, *39*(5), 550–557.

[47] **Liao, S., van Delft, C. and Vial, J.P.** (2013). Distributionally robust workforce scheduling in call centres with uncertain arrival rates, *Optimization Methods and Software*, *28*(3), 501–522.

[48] **Rohleder, T., Bailey, B., Crum, B., Faber, T., Johnson, B., Montgomery, L. and Pringnitz, R.** (2013). Improving a patient appointment call center at Mayo Clinic, *International Journal of Health Care Quality Assurance*, *26*(8), 714–728.

[49] **Chan, W., Koole, G. and l'Ecuyer, P.** (2014). Dynamic call center routing policies using call waiting and agent idle times, *Manufacturing & Service Operations Management*, *16*(4), 544–560.

[50] **Peng, Y., Qu, X. and Shi, J.** (2014). A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients, *Computers & Industrial Engineering*, *72*, 282–296.

[51] **Defraeye, M. and Van Nieuwenhuyse, I.** (2016). Staffing and scheduling under nonstationary demand for service: A literature review, *Omega*, *58*, 4–25.

[52] **Aktekin, T. and Ekin, T.** (2016). Stochastic call center staffing with uncertain arrival, service and abandonment rates: A Bayesian perspective, *Naval Research Logistics (NRL)*, *63*(6), 460–478.

[53] **Andersen, A.R.**, **Nielsen, B.F.**, **Reinhardt, L.B. and Stidsen, T.R.** (2019). Staff optimization for time-dependent acute patient flow, *European Journal of Operational Research*, *272*(1), 94–105.

[54] **Shen, H. and Huang, J.Z.** (2008). Forecasting time series of inhomogeneous Poisson processes with application to call center workforce management.

[55] **Ding, S.**, **Koole, G. and van der Mei, R.D.** (2015). On the estimation of the true demand in call centers with redials and reconnects, *European Journal of Operational Research*, *246*(1), 250–262.

[56] **Avramidis, A.N.**, **Deslauriers, A. and L'Ecuyer, P.** (2004). Modeling daily arrivals to a telephone call center, *Management Science*, *50*(7), 896–908.

[57] **Parisio, A. and Jones, C.N.** (2015). A two-stage stochastic programming approach to employee scheduling in retail outlets with uncertain demand, *Omega*, *53*, 97–103.

[58] **Ta, T.A.**, **Chan, W.**, **Bastin, F. and L'Ecuyer, P.** (2021). A simulation-based decomposition approach for two-stage staffing optimization in call centers under arrival rate uncertainty, *European Journal of Operational Research*, *293*(3), 966–979.

[59] **Banks, J.** (2005). *Discrete event system simulation*, Pearson Education India.

[60] **Kelton, W.D.**, **Sadowski, R.P. and Sadowski, D.A.** (2002). *Simulation with ARENA*, McGraw-Hill, Inc.

[61] **Choi, B. and Kang, D.** (2013). *Modeling and Simulation of Discrete-Event Systems*.

[62] **Altiok, T. and Melamed, B.** (2007). *Simulation Modeling and Analysis with ARENA*.

[63] **Rossetti, M.** (2015). *Simulation Modeling and Arena*, Wiley Series in Modeling and Simulation Series, Wiley, `https://books.google.com.tr/books?id=j2xuCAAAQBAJ`.

APPENDICES

APPENDIX A : Tables

Table A.1 : Classification of literature studies on staff scheduling

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Ma, J., Kim, N., & Rothrock, L. [27]		X			Discrete-Event Simulation	
Dietz, D. C. [46]	X	X		Quadratic Programming	Markovian Queuing	
Taylor, J. W. [29]			X			Holt-Winters Exponential Smoothing
Liao, S., van Delft, C., & Vial, J. P. [47]	X	X	X	Stochastic Programming	Monte-Carlo Simulation	
Van den Bergh, J., Beliën, J., De Bruecker, P., Demeulemeester, E., & De Boeck, L. [9]	X			Metaheuristics		

Table A.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Rohleder, T., Bailey, B., Crum, B., Faber, T., Johnson, B., Montgomery, L., & Pringnitz, R. [48]	X	X		Tabu-Search Algorithm	Discrete-Event Simulation	
Gabrel, V., Murat, C., & Thiele, A. [10]	X			Robust Stochastic Optimization		
Chan, W., Koole, G., & L'Ecuyer, P. [49]	X	X		Modified Genetic Algorithm		
Peng, Y., Qu, X., & Shi, J. [50]	X	X		Genetic Algorithm	Discrete-Event Simulation	
Lin, S. W., & Ying, K. C. [11]	X			Three-Phased Iterated Greedy Heuristic		

Table A.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Parisio, A., & Jones, C. N. [57]	X	X	X	Two-stage Stochastic Optimization (Mixed-Integer Linear Programming)		Least-Squares Method
Ding, S., Koole, G., & van der Mei, R. D. [55]		X	X		Discrete-Event Simulation	Auto-Regressive Integrated Moving Average
Excoffier, M., Gicquel, C., & Jouini, O. [12]	X			Mixed-Integer Linear Programming & Stochastic Programming		
Walter, M., & Zimmermann, J. [13]	X			Mixed-Integer Linear Programming		

Table A.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Defraeye, M., & Van Nieuwenhuyse, I. [51]	X	X		Stochastic Programming	Discrete-Event Simulation	
Aktekin, T., & Ekin, T. [52]	X	X		Stochastic Programming	Augmented Probability Simulation	
Ibrahim, R., Ye, H., L'Ecuyer, P., & Shen, H. [30]			X			Time-Series Forecasting
Taskiran, G. K., & Zhang, X. [14]	X			Mixed-Integer Programming		
Mattia, S., Rossi, F., Servilio, M., & Smriglio, S. [15]	X			Two-Stage Integer Programming		
Bodur, M., & Luedtke, J. R. [16]	X			Stochastic Integer Programming		

Table A.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Tang, X., Liao, X., Zheng, J., & Yang, X. [31]			X			Wawelet Neural Network & Multilinear Regression
Vermuyten, H., Rosa, J. N., Marques, I., Belien, J., & Barbosa-Póvoa, A. [17]	X			Integer Programming& VNDS heuristic& Dividing Heuristic		
Andersen, A. R., Nielsen, B. F., Reinhardt, L. B., & Stidsen, T. R. [53]	X	X		Integer Programming & Recursive Bound Adaptation Meta-heuristic	Discrete Event Simulation	

Table A.1 (continued) : Classification of literature studies on staff scheduling.

Study	Optimization	Simulation	Forecasting	Type of Optimization	Type of Simulation	Type of Forecasting
Ta, T. A., Chan, W., Bastin, F., & L'Ecuyer, P. [58]	X	X		Two-Stage Stochastic Optimization	Several Independent Simulations	
Xu, Y., & Wang, X. [18]	X			Artificial Bee Colony Algorithm		
Boysen, N., Emde, S., & Schwerdfeger, S. [20]	X			Bender's Decomposition		
Mansini, R., Zanella, M., & Zanotti, R. [21]	X			Mixed-Integer Linear Programming		
Chacón, H., Koppisetti, V., Hardage, D., Choo, K. K. R., & Rad, P. [32]			X			Holt-Winters Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average

CURRICULUM VITAE

Name SURNAME: Ömer Faruk Kandaz

EDUCATION:

- **B.Sc.:** 2020, Istanbul Technical University, Faculty of Civil Engineering, Civil Engineering Department
- **M.Sc.:** 2024, Istanbul Technical University, Faculty of Management, Management Engineering Department

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2020 Completed Bachelor's at İstanbul Technical University.
- 2024 Completed Master's at İstanbul Technical University.
- 2022-Present Research Assistant at Istanbul Technical University.

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- **Kandaz O.F.**, **Ozcan T.** (2024). A Simulation-Based Staff Scheduling Analysis for Call Centers in Cargo Industry. *International Conference on Intelligent and Fuzzy Systems*, July 16-18, 2024 Canakkale, Turkey. (Presentation Instance)

OTHER PUBLICATIONS, PRESENTATIONS AND PATENTS:

- **Alper T.**, **Kandaz Ö.F.** & **Akdağ H.C.** (2023). Application of Sustainable Lean Supply Chain Management for a Textile Company. *The International Symposium for Production Research*, October, 2023 Antalya, Turkey. (Conference Paper)
- **Kandaz Ö.F.**, **Akdağ H.C.** (2023). The Relationship Among Sustainable Development, GDP and Greenhouse Gas Emission in Turkey. *The International Symposium for Production Research*, October, 2023 Antalya, Turkey. (Conference Paper)