İSTANBUL UNIVERSITY PRESS

Istanbul Business Research, 51(1)
DOI: 10.26650/ibr.2022.51.951646
http://ibr.istanbul.edu.tr/
http://dergipark.org.tr/ibr

Istanbul Business Research

Submitted: 14.06.2021
Revision Requested: 26.09.2021
Last Revision Received: 02.10.2021
Accepted: 01.11.2021
Published Online: 26.01.2022

RESEARCH ARTICLE

Discrete Event Simulation Model Performed with Data Analytics for a Call Center Optimization

Nisan Güniz Serper[1], Elif Şen[2], Banu Çalış Uslu[3]

Abstract

Optimization models enable organizations to find the best solution and respond to the demand from an uncertain environment and stochastic process promptly and with less engineering effort. This study aims to optimize the number of seasonal agents and customer prioritization needed for a call center system using big data analytics and discrete event simulations to improve customer satisfaction. The study was carried out based on data from a leading heating and ventilation company's call center. The K-means clustering technique was used to determine customer segmentation on 6-million-customer data. For prioritization, the making of a Recency-Frequency-Monetary (RFM) analysis was applied. The system was modeled using ARENA simulation software, and performance parameters were measured depending on the segments obtained. The results show that the simulation model performed with data analytics gives better results for a beneficial financial impact with numerical values in customer prioritization, reducing the average waiting time of the most prioritized customers by more than 90%, and for the least prioritized customers, it increased the average waiting time by approximately just 40%. However, with the company segments, the increase in the average waiting time of the least prioritized customers was approximately 300%.

Keywords

Call Center Management, Simulation, Prioritization, Data Analytics, Customer Segmentation

Introduction

Call centers have become the heart of the service sector, which aims to manage customer interactions to sustain operations (Ma, J. et al., 2011). It serves a function in determining customer loyalty and satisfaction with a company. It is strategically vital (Anton, 2000) as call center management is a communication channel where companies directly contact customers (K.J. et al., 2004; Saberi et al., 2017). To effectively and efficiently manage the call centers, companies must carefully analyze the relevant parameters, processes, logic, and relationships to perform effective management. (Mehrotra and Fama, 2003; Feinberg et al., 2000). The process is that customers call the call centers, ask questions about their challenges, compla-

1 Nisan Güniz Serper (Student), Marmara University, Faculty of Engineering, Department of Industrial Engineering, Istanbul, Turkey. E-mail: nisanserper@gmail.com ORCID: 0000-0001-8981-3048

2 Elif Şen (Student), Marmara University, Faculty of Engineering, Department of Industrial Engineering, Istanbul, Turkey. E-mail: elifsenie@gmail.com ORCID: 0000-0002-0056-3204

3 **Corresponding Author:** Banu Çalış Uslu (Dr.), Marmara University, Faculty of Engineering, Department of Industrial Engineering, Istanbul, Turkey. E-mail: bcalis@marmara.edu.tr ORCID: 0000-0001-8214-825X

To cite this article: Serper, N. G., Sen, E., & Calis Uslu, B. (2021). Discrete Event Simulation Model Performed with Data Analytics for a Call Center Optimization. *Istanbul Business Research, 51*(1) Advanced Online Publication. http://doi.org/10.26650/ibr.2022.51.951646

ints, or the needed solutions for their problems. The company must attach importance to call centers because if customers are not satisfied, there might be a direct loss for a business, i.e., unsatisfied customers may not make repurchases (Feinberg et al., 2000).

Data Analytics is commonly used to analyze and understand customer data. Customer segmentation is an essential technique to gain insight on customer behaviors for decision making and developing strategy. Segmentation divides customers into discrete, homogenous customer groups based on similar characteristics or purchasing preferences. (Hassan, M. M., and Tabasum, M., 2018)

Discrete-Event Simulation (DES), which is one of the strategic management tools, is useful in decision-making by looking at problems as a whole and expressing all the relationships, interactions, and uncertainty sets (Arora, 2007). The simulation is appropriate for analyzing the relationships which are significant for call center management, such as waiting time in queues, source management, utilization, etc. Since DES is very popular in assessing and optimizing stochastic processes, a wide variety of applications exist in the literature. Some of them are as follows: optimization of raw material allocation process (Windisch et al., 2015), increasing operational and manufacturing efficiency (Troncoso-Palacio et al., 2018; Riskadayanti et al., 2019; Knapčíková et al., 2020), optimization of transportation problems (Afrapoli et al., 2019; Behiri et al., 2018), line balancing (Bongomin et al., 2020; Yemane et al., 2020), and stock management (Gittins et al., 2020; Amorim-Lopes et al. 2021).

This study developed a DES model with data analytics to increase customer satisfaction by determining the number of agents needed and the most appropriate segment for a real-life call center. To the best of our knowledge, this is the first study to assess call center performance management by using simulation and data mining in the literature. The proposed model is believed to be a guide for call center managers and performance management professionals.

The call center of one of the leading companies in the heating, cooling, and air-conditioning sector was selected as a case study. This company receives almost 3.5 million calls during the year from among their 6-million customers. Due to long waiting times, we constructed data-driven customer segmentation applying k-means clustering on the 6-million-customer data and the prioritization making recency-frequency-monetary (RFM) analysis. Then, we compared data-driven segments with the current company segments using a simulation methodology. As a result of the study, the proposed model makes the company serve valuable customers with a waiting time of fewer than 5 seconds, reducing the number of agents by 15%, saving labor costs, and making data-driven customer segmentation prioritization. The research conducted within this research scope clearly shows that big data methods give better results than expert approaches in prioritizing customers. For this reason, extensive data analysis will be inevitable before the simulation or any optimization method to be implemented.

The research methodology of this study is represented in Figure 1. The literature review includes current studies in call center management through google scholar and web of science. Then, the data analysis phase is processed for input data such as interarrival time and process time. For seasonality and dependency, time series graphs were created, and autocorrelation tests were made from the data, respectively. Next, K-Means and clustering were applied to the customer data for segmentation after k was defined by using the Elbow Method. Then, the RFM Method was used to prioritize the clusters which were a result of clustering. Another important step was DES Modelling. In this step, firstly, suitable distributions were determined through the input analyzer, homogeneity tests were checked, and then, the process was modeled via ARENA simulation software. After the verification and validation tests on the developed model were successful, the scenario analysis step was started. Finally, the scenario analysis was done based on the performance parameters, and then, alternative scenarios were compared.

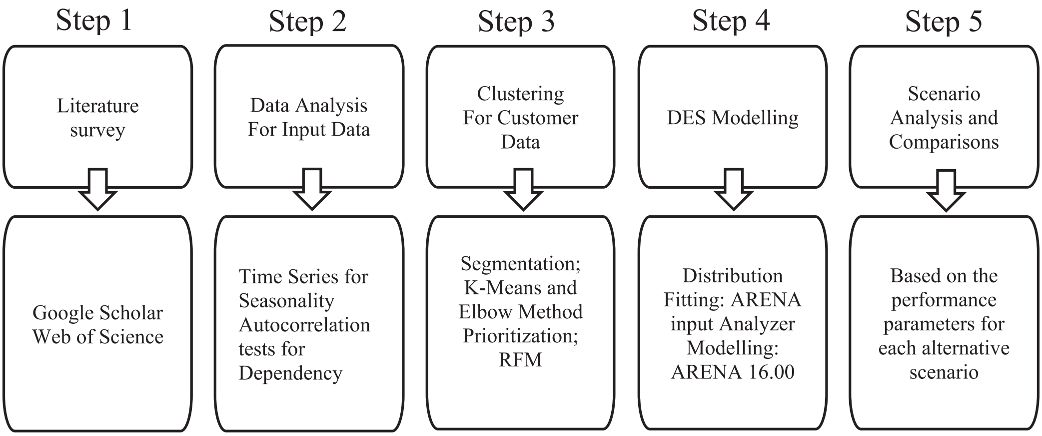


Figure 1. Research Methodology Diagram

The rest of the paper is organized following the strengthening of the reporting of empirical simulation studies (STRESS) guidelines (Monks *et al.*, 2019). STRESS guidelines include the *literature review section*, which demonstrates the current applications of call center management in the literature; the *objectives* section, which explains the purpose and the expectations of the study; the *logic* and *scenario logic* sections, which contain the current and scenario logics and explain the differences between the two models; the *data* section, which contains parameters and assumptions for the model; the *experimentation* section, which shows the initialization parameters of the simulation model; the *implementation* section, which shows the technical background for the running of the simulation model; the *result* section, which contains the comparison of model outputs; and the *conclusion* section, which includes the inferences made by the study, respectively (Monks *et al.*, 2019).

Literature Review

In the literature, there are several studies related to performance management at call centers. Legros et al. (2017) enhanced a Markov chain method that evaluated a call center model's performance, which provides a feature of converting an incoming call to an outgoing call for performance evaluation and queue optimization (Legros et al., 2017). Van Buuren et al. (2017) investigated and equated three discrete-event simulation models for emergency department call centers and then obtained some inferences on how to decrease response time. Simulation is a convenient way to evaluate a call center system's performance since it allows examining the different scenarios and can be combined with data analysis (Lam and Lau, 2004; Doomun and Jungum, 2008).

Ibrahim *et al.* (2016) reviewed the literature on forecasting and modeling call arrivals, discussed the critical points for structuring a qualified statistical arrival model, and measured the proportions of forecasting precision. These values were obtained from call center data, which was collected in real life. Moreover, they stated that to achieve better forecasting accuracy and effective operational decisions, call center arrivals to play an important role were based on scheduling, routing, and staffing. Aktekin (2014) studied how to reduce the paucity of studies on the uncertainty of input distributions and their impact on call center management. Then, he observed that different customer profiles require other agent skills. Thus, the anticipation of service distribution may decrease.

Alotaibi and Liu (2013) developed a new numerical model to improve the average waiting times of a telecommunications contact center by prioritizing customer groups. The authors primarily aimed to improve the customer satisfaction of customer groups with high priority, and they optimized the waiting times of prioritized groups using the proposed model. Abdullateef *et al.* (2010) proposed a conceptual framework for a customer contact center. They aimed to evaluate the critical factors of sufficient caller satisfaction and service delivery. They stated that, in addition to a well-performed first call experience, the four primary elements (client, technology, process, and people) have a crucial role in caller satisfaction. Koole and Pot (2005) committed a model that assigning jobs to employees in multi-skill call centers often occurs according to priority routing policies in conjunction with agent groups. They observed that, compared to the initial model, assigning calls to specialists reduces the average queue length by 50 percent.

To understand customer behavior, customer profiles should be determined (Farruh, 2019; Anshari et al., 2019, Thomas and Shirani, 2020; Greco and Polli, 2020; Gayathri et al., 2020). Determining customer profiles includes five models: segmentation, customer profitability, customer retention, customer clustering, and response analysis (Hahnke, 2001). Customer segmentation is an essential technique to understand customer behavior for decision making and developing strategy. Segmentation divides customers into discrete, homogenous customer groups based on the similarity of characteristics or purchasing preferences (Hassan, M. M., and Tabasum, M., 2018; Carnein and Trautmann, 2019; Hung et al., 2019). Customer clustering finds similarities/relationships between data sets using data mining techniques and uses those similarities to create meaningful clusters (Saglam et al., 2006; Shih et al., 2010; Rudskaia and Eremenko, 2019). Clustering analysis divides the data records into classes, but the same class data are very similar while the data in different classes are quite distinct (Klement, P., and Snášel, V., 2011; Jintana and Mori, 2019).

In addition to the importance of understanding the customer for call center performance, simulating the process of a call center is another critical point to see the current performance and improve it. Avramidis and L'Ecuyer (2005) mentioned the importance of call centers for companies and explained why call centers need simulations. The main reasons were the complexity of call center problems and a simulation's capability of solving complex problems straightforwardly (Calis, 2016). It was observed in this study that simulations benefited companies as follows: Companies can see the effects of the changes they plan to make without interrupting operations. The call center is finally emerging as a manageable, responsive, and customizable strategic weapon with simulation. Mehrotra and Fama (2003) studied how call centers use simulations to overview call center simulation models, emphasizing characteristic inputs and data sources, modeling challenges, and critical model outputs. They created a simulation model and determined routing strategies for a call center using simulation results obtained from 3 different agent groups.

Objectives

The study was conducted based on the data of a leading heating and ventilation company's call center. The company is a leading company in international markets for heating and ventilation technology and serves Turkey within regional offices with authorized dealers and service technicians. The company's call center receives almost 3.5M calls during the year. This study aims to propose a model to increase customer satisfaction through the determination of the most suitable segment and needed number of agents for the system.

Logic

In this part of the research, the current system analysis and segments defined by the company are presented. Then, the results obtained from big data analytics and the analyses made in this context are explained. Finally, the findings obtained by comparative simulation analyzes are shared.

Base Model Overview

In the company's current model, the customer arrives and directly goes to the interactive voice response (IVR) system, which is an automatic speech system that is used to orient customers for the process (Dillman *et al.*, 2009). In the company's call center, every customer is delayed on IVR for 47 seconds. After IVR, some calls miss, and some of them go to the call process. Then, the customer leaves. Segments defined by the company are given in Table 1

Table 1
Customer segments determined by the company

Segment Name	**Rules**	**Call Priority**
S1: Exclusive Device Customer	Customers who have bought determined prioritized hero products.	1
S2: Contracted Loyal Customer	Customers who have purchased a maintenance contract for at least six years at one time or during each renewal period.	2
S3: Contracted Customer	Customers who have purchased a maintenance contract for one year or have purchased a product with no warranty agreement and receive maintenance or breakdown service for three consecutive years.	3
S4: Inactive Customer	Customers who purchased a product more than five years ago with no purchase in the past five years.	4

Data Mining and Customer Segmentation

With technological developments and digital transformation in this age, data grows, and the importance of using data correctly increases (David, 2013; Verhoef *et al.*, 2019). As data grows, it becomes more complex and incomprehensible (Erevelles *et al.*, 2016; Şen et al., 2019; Uslu & Fırat, 2019). Companies that cannot improve toward analyzing and using big data may have difficulty keeping up with the trend in terms of competition and performance (Watson, 2012; Uslu, 2020).

The research aims to form customer groups based on their characteristics to prioritize them in the call center. This process is called customer segmentation, which is defined by Tsiptsis and Chorianopoulos (2011) as a division process of customers into homogeneous groups based on their behaviors. Since firms have recently obtained vast customer data, data mining is becoming an appropriate way to analyze it (Rygielski *et al.*, 2002).

In the literature, several studies about customer segmentation exist. When those studies are examined in detail, it can be seen that some researchers analyze behavioral features of clusters after clustering. Rajagopal (2011) used customer clustering to identify five customer groups using a determined number of clusters. After the author determined the clusters, he analyzed their profit profiles using Search Query Language (SQL) to see their lifetime value. Farajian and Mohammadi (2010) analyzed the attitudes of banking customers using clustering and association rule techniques. In the clustering section, first, they applied K-means clustering, and they supposed that k equaled 3. Then, they calculated Recency, Frequency, and Monetary (RFM) scores to reach the meanings of clusters.

Additionally, clustering is applied after analyzing behaviors using some specific techniques by some researchers. Shih and Liu (2003) made K-means clustering based on RFM weights, and they determined k as 8. Then, they ranked the clusters in terms of customer lifetime value. Namvar et al. (2010) made customer clustering using demographic variables on the data of Iranian bank customers. First, they calculated the RFM scores of each customer, and then, they used K-means clustering with 9 clusters. In this project, first, segmentation would be made based on customer transaction data using K-means clustering in R programming language; then, the prioritization would be made using RFM Analysis.

For the customer segmentation, the disinfected customer data taken from the company contained 6,030,355 customers, and 16 attributes of each customer were used applying k-means clustering and RFM analysis. K-means methodology is one of the most common methods used for customer clustering (Figueiredo *et al.*, 2003; Niyagas *et al.*, 2006; Windorto *et al.*, 2019; Maheshwari *et al.*, 2019; Rojlertjanya, 2019; Gustriansyah *et al.*, 2020; Mousavi *et al.*, 2020; Nugraha, 2020). The primary purpose of the k-means clustering is to form clusters that "minimize the squared error criterion" using the predetermined number of k values, which represents the number of clusters (Ye *et al.*, 2013). To obtain an optimal number of clusters, the Elbow Method's interpretation would be appropriate before applying k-means clustering (Bholowalia and Kumar, 2014; Syakur *et al.*, 2018; Anuşlu and Fırat, 2019; Nainggolan *et al.*, 2019; Cui, 2020; Liu and Deng, 2020; Umargono *et al.*, 2020).

For the clustering phase, the following variables were used: number of products, agreement contracts, number of maintenance contracts, and duration of agreement contracts. First, the Elbow method was applied, and the optimal value of k was found as 8. Then, using the optimum k-value and k-means clustering, data-driven segments were obtained.

For segment prioritization, RFM analysis was made based on customer transactions. During the RFM application, first, *recency value* as the time between the last transaction and the present, *frequency value* as the number of transactions, and *monetary value* as the price of the transactions were calculated for each customer in each cluster based on products and contracts (Zalaghi and Varzi, 2014; Kadir and Achyar, 2019; Maraghi *et al.*, 2020). All criteria scores were then combined and ranked to form an overall RFM score between 1 and 5 (Zalaghi and Varzi, 2014; Sabuncu *et al.*, 2020). Second, according to the average RFM scores, clusters were prioritized as seen in Table 2.

Table 2
The segments which were determined by the clustering method and their calculated RFM scores

Clusters	Rules	RFM Score
Cluster 6	Those have at least one combi, thermosiphon, and agreement contract.	2.996
Cluster 2	Those have not heat pump and/or just have a boiler but have not agreement contract.	1.930
Cluster 4	Those have not heat pump and cascade but have at least one combi or thermosiphon.	1.743
Cluster 8	Those do not have to cascade but have at least two air conditioners.	1.666
Cluster 5	Those have not heat pump and cascade but have thermosiphon.	1.569
Cluster 7	Those have not cascade and agreement contracts but have thermosiphon.	1.390
Cluster 1	Those have not to a heat pump but have the geyser.	1.212
Cluster 3	Those have not to heat pump and cascade but have at least one combi or geyser	1.197

Time Series Analysis and Homogeneity Test

To check the seasonality effect, time-series graphs were formed based on weekly average process times and interarrival times, and the process was considered within four seasons to prevent seasonality (Nwogu *et al.*, 2016). As seen in Figure 2, process times have seasonality. After it was decided to proceed by dividing the data into four seasons to avoid seasonality, autocorrelation graphs were formed with RStudio to measure the dependency in the data (Banks et al, 2005). As can be seen in the example graph for process times of season three in Figure 3, since there were some lines that went beyond the upper line of the graph, it was necessary to analyze the residuals to check whether the residuals had trends or not. Using Excel, residual graphs were formed for both interarrival and process times of each season. According to the residual graphs, all the data points of each season lie flat, so there was no trend on the residual graphs. An example residual graph of the process time of season three can be seen in Figure 4 below. Then, autocorrelation graphs were formed to check data dependency (Banks *et al.*, 2005). Some lines go beyond the graph's upper line, so it was necessary to analyze the residuals to check whether the residuals had trends. Residual plots were formed for each season's interarrival and process times, and there was no trend on the residual graphs.

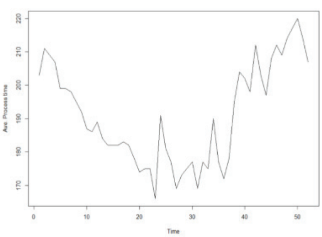


Figure 2. Time series graph for process time

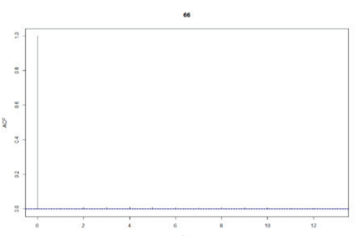


Figure 3. An example autocorrelation graph for process times of Season 3

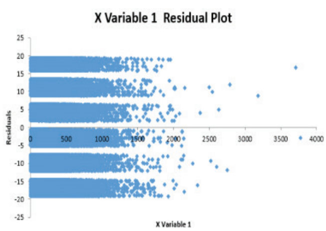


Figure 4. Example residual graph for process times of Season 3

After autocorrelation, using the Input Analyzer tool of ARENA, probability distributions of interarrival times and process times for each season were obtained, as shown in Table V. To decide which distribution had a good fit, the corresponding p-value of the Chi-square test was considered. (Banks *et al.*, 2005; Andrade, 2019).

Resources

Agents are the call center representatives who communicate with customers. In the call center, they are responsible for making inbound and outbound calls. According to the company's information, the ratio of agents making inbound calls changes from season to season depending on the traffic of calls. In the company, 1559 agents are working both in-source and out-source within shifts of 11 hours.

Verification and Validation

Verification is about building the model correctly while validation is about building the correct model (Sargent, 2013). That is why, for verification, a checklist that includes all elements of the model was prepared, and it was asked by an expert from the company whether all the system parameters were considered in the model or not.

Additionally, validation is a statistical analysis that compares observations with the simulation model results (Banks *et al.*, 2005). To validate the simulated model, hypothesis testing was used, and α was selected as 0.01 (Banks *et al.*, 2005). As a result, the model is found valid upon these statistical calculations.

Scenario Logic

There were two scenarios experienced, so the first scenario was prioritization according to the segments determined by the company whereas the second scenario was prioritization according to the segments determined by the clustering method.

First Scenario: Prioritization according to the Company Segments

Only the resource's queue type and schedule were changed to build an alternative scenario in the first scenario. Queue type was first-in, first-out in the current model, but the queue was prioritized according to the first alternative scenario segments. The schedule of resources means that the number of agents was reduced at a rate of 15%. The distribution of customer types is shown in Table 3. These ratios were used for customer arrivals of the call center system.

Second Scenario: Prioritization according to Proposed Data Analytics Segments

In the second scenario, only the queue type and schedule of the resources were changed to build an alternative scenario. Queue type was first-in, first-out in the current model, but the queue was prioritized according to the segments in the second alternative scenario. The schedule of resources means that the number of agents was reduced at a rate of 15%. The eight clusters were obtained by using the clustering method. Then, they were segmented by using the RFM method, and ordered segments were obtained. The distribution of these segments are shown in Table 3 and Table 4. The ratios seen below were used for customer arrivals of the call center system

Table 3
Distribution of calls by company customer segments for each season

	Segment	Percentage of calls from the segment		Segment	Percentage of calls from the segment
SN1	S1	2,80%	SN3	S1	2.5%
	S2	5.5%		S2	4.9%
	S3	75.2%		S3	76.8%
	S4	16.5%		S4	15.8%
	Segment	**Percentage of calls from the segment**		**Segment**	**Percentage of calls from the segment**
SN2	S1	2.4%	SN4	S1	2.2%
	S2	7.2%		S2	4.3%
	S3	75.0%		S3	76.0%
	S4	15.4%		S4	17.5%

Table 4
Distribution of calls by clustering segments for each season

	Segment	Percentage of calls from the segment		Segment	Percentage of calls from the segment
SN1	Cluster 6	0.1%	SN3	Cluster 6	0.1%
	Cluster 2	0.8%		Cluster 2	1.0%
	Cluster 4	0.8%		Cluster 4	0.8%
	Cluster 8	2.7%		Cluster 8	3.2%
	Cluster 5	8.4%		Cluster 5	10.4%
	Cluster 7	7.4%		Cluster 7	7.2%
	Cluster 1	7.4%		Cluster 1	8.3%
	Cluster 3	72.3%		Cluster 3	69.0%
	Segment	**Percentage of calls from the segment**		**Segment**	**Percentage of calls from the segment**
SN2	Cluster 6	0.1%	SN4	Cluster 6	0.2%
	Cluster 2	0.9%		Cluster 2	0.7%
	Cluster 4	0.6%		Cluster 4	0.8%
	Cluster 8	2.8%		Cluster 8	2.5%
	Cluster 5	10.2%		Cluster 5	8.4%
	Cluster 7	7.0%		Cluster 7	8.1%
	Cluster 1	8.6%		Cluster 1	7.1%
	Cluster 3	69.8%		Cluster 3	72.2%

Data Analysis

There are two different data sets: the device purchases of 6M customers and call center data for one year. Both of the data sets were taken from the company as cleaned, so there was no need to do any pre-processing. The call center data was broken into four seasons to eliminate seasonality effects. The missing call percentage of the system was calculated based on historical data of each season, as the share of missing customers overall. The data set contained arrival time and service time. To illustrate the model of the call center's current system, input parameters and their distributions are shown in Table 5. The ARENA Input Analyzer was used to determine appropriate distributions. To decide which distribution had good fit, the corresponding p-value of Chi-square test was considered. If the p-value was less than 0.05, it could be said that the distribution was fit for the data (Banks *et al.*, 2005; Andrade, 2019).

Table 5
Input Parameters of Current Model

Season	Interarrival Time Distribution	Process Time Distribution	Missing Call Percentage
S1	-0.001 + EXPO(7.33)	0.999 + GAMM(109, 1.7)	1.3%
S2	-0.001 + WEIB(9.84,0.783)	0.999 + GAMM102, 1.65)	1.2%
S3	-0.001 + EXPO(9.14)	0.999 + GAMM(106, 1.62)	1.4%
S4	-0.001 + EXPO(4.69)	0.999 + GAMM(112, 1.75)	1.7%

To develop the model, some assumptions were made, and these can be seen in Table 6.

Table 6
Assumptions of the current model of the call center of the company

Assumptions
• The capacity of the line is limitless.
• There is no breakaway in the line.
• The call center serves 11 hours a day from 08:00 am to 07:00 pm, and only on weekdays.
• Distributions of process and interarrival times are the same for the current model and alternative models.
• There is no difference between agents in terms of performance.

Experimentation

Both models, current and scenario, had a run length of 90 days for each season. The number of initial replications was determined as five, and there was no warm-up period in both models. No more than five replications were applied because the model was validated. (Banks *et al.*, 2015). All of the estimations were based on an average of 5 replications of each model for each season. Working time was 11 hours per day, five days a week from Monday to Friday. In the beginning, there was no queue in the system. The first call came on the first day, at time zero.

Implementation

Both models were implemented by Arena Simulation Software 16.00.00000 full version on Lenovo Yoga with 6. Generation Intel® Core™ i7 CPU and 8 GB RAM. The average runtime of each season data based on the five replications is given in Tables VII and VIII.

Results

Based on three performance parameters, this part of the study compares three different simulation models to identify the best scenario to increase call center system performance.

These parameters are:

1. Comparison of average waiting times are given in Table 7 and Table 8

2. Comparison of the number of customers by different waiting time scale is given in Table 9

3. Comparison of agent utilization of the current system and agent utilization after reducing the number of agents by 15% is given in Table 10

Table 7
Waiting times of company segments for each season

	Average Waiting Time (sec)		
	Segment Name	**Average**	**Runtime**
Season 1	S1	1.22	00:35:37
	S2	1.35	
	S3	**18.13**	
	S4	**192.69**	
Season 2	S1	3.01	00:53:24
	S2	3.37	
	S3	**22.98**	
	S4	**274.35**	
Season 3	S1	2.57	00:21:50
	S2	2.73	
	S3	**25.93**	
	S4	**309.66**	
Season 4	S1	2.03	01:02:48
	S2	2.15	
	S3	**22.52**	
	S4	**741.22**	

Table 8
Waiting times of clustering segments for each season

	Average Waiting Time (sec)		
	Segment Name	**Average**	**Runtime**
Season 1	C6	1.082	00:49:37
	C2	1.172	
	C4	1.228	
	C8	1.292	
	C5	1.438	
	C7	1.766	
	C1	**2.162**	
	C3	**62.198**	
Season 3	C6	2.79	00:24:53
	C2	2.90	
	C4	3.08	
	C8	3.14	
	C5	3.69	
	C7	4.44	
	C1	**5.49**	
	C3	**83.73**	
Season 1	C6	2.506	00:44:16
	C2	2.528	
	C4	2.58	
	C8	2.654	
	C5	3.106	
	C7	3.744	
	C1	**4.622**	
	C3	**98.588**	
Season 3	C6	2.01	01:23:11
	C2	2.02	
	C4	2.02	
	C8	2.10	
	C5	2.37	
	C7	2.88	
	C1	**3.52**	
	C3	**202.37**	

Comparisons of Current and Proposed Models

For each season and each segment of both from the company and clustering in the study, a simulation was run with the FIFO and a prioritization. In conclusion, it can be said that there was ***a reduction with a minimum rate of 90%*** on the segments excluding the last ones for both types of segmentation (Please see the Table 7 and Table 8).

Comparison of Average Waiting Times of Company Segments and Clustering Segments

One of the significant performance measures of customer satisfaction of a call center is response time to a customer call (Robinson and Morley, 2006). In this section, a comparison was made based on the number of customers having specific waiting times. Table 9 shows the corresponding number of customers based on waiting time intervals in each season.

When we compare company segments and cluster segments by waiting time, we can see that cluster segments were more valuable to satisfy customers because customers prioritized according to ***cluster segments were waiting less than 5 seconds.*** Also, company segment prioritization shows that many customers waited for more than 100 seconds in all seasons except the fourth season.

Table 9
Comparison of the number of customers by different waiting time scale

	Season 1			Season 2	
Waiting Time	**Company Segment**	**Cluster Segments**	**Waiting Time**	**Company Segment**	**Cluster Segments**
0-5 sec	1371	4568	0-5 sec	962	2172
5-25 sec	12407	0	5-25 sec	7511	857
25-65 sec	0	11938	25-65 sec	0	0
65-100 sec	0	0	65-100 sec	0	6991
100+ sec	2723	0	100+ sec	1547	0
	Season 3			**Season 4**	
Waiting Time	**Company Segment**	**Cluster Segments**	**Waiting Time**	**Company Segment**	**Cluster Segments**
0-5 sec	794	3317	0-5 sec	1077	4582
5-25 sec	0	0	5-25 sec	12519	0
25-65 sec	8221	0	25-65 sec	0	0
65-100 sec	0	7391	65-100 sec	0	0
100+ sec	1694	0	100+ sec	2877	11891

Resource Utilization of Call Center Agents

Finally, the last comparison is about resource utilization of call center agents. While running simulation models of alternatives, we reduced the number of workers at a rate of 15%. The following table shows the utilization rates of each season before reduction and after reduction.

Table 10
Utilization rates of each season before reduction and after reduction

Season	Utilization of Current System	Utilization after Reduction
S1	53%	**62%**
S2	62%	**72%**
S3	61%	**72%**
S4	71%	**82%**

As shown in Table 10, the company can save labor costs and increase utilization by reducing the number of employees. Moreover, despite this reduction, it can also have efficient waiting times.

Conclusion

In conclusion, three main findings were obtained. Firstly, applying data mining techniques reduced the average waiting time of most prioritized customers by more than 90%, and for the least prioritized customers, it increased the average waiting time just by approximately 40%. However, with the company segments, the increase in the average waiting time of the least prioritized customers was approximately 300%. Secondly, as a result of the comparison of the number of customers by different waiting time scale, it was seen that customers were waiting less than 5 seconds within the segmentation by big data analysis. Finally, resource utilization was increased by reducing the number of agents with a rate of 15%.

Going forward, two main trends affecting call center simulation were seen. Firstly, the operational complexity will continue to grow with the increase of digitalization. As a result of the rise in outbound services daily, more hybrid methodologies will need to be created, and more queues will occur. Better optimization models for the management of agents will need to be designed to manage call center systems. This will create a significant challenge in defining metrics to be developed for management. The simulation models to be developed not only for the management of incoming calls but also for managing total call volumes, system capacity, and customer satisfaction will also need to include risk analysis and experimental design techniques.

Secondly, with the increasing use of big data globally, big data should be examined with more effective models for developing call center systems. The research conducted within this study clearly shows that big data methods give better results than expert decisions in prioritizing customers. For this reason, data mining will be inevitable before the simulation or any optimization method are implemented.

The key limitation of the study is that it is not suitable for night shifts due to the high variability of input data. In addition to this, each agent's performance is accepted as equal because the case study company did not provide the agents' performance data since the data cannot be anonymized based on Turkish Personal Data Protection Law. Therefore, it needs to be adjusted for different capabilities where different levels of competence are essential.

The Covid-19 epidemic process showed that the importance of call centers' effective use is increasing day by day. In particular, call centers played an active role in resolving the problems of customers who could not reach the seller directly during this process. It is planned to develop the research in this direction, examine the effect of the Covid-19 outbreak on the call center's process performance, and develop a risk management model in this context.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of study: B.Ç.U., E.Ş., N.G.S.; Data Acquisition: N.G.S..; Data Analysis/Interpretation: E.Ş., N.G.S., B.Ç.U.; Drafting Manuscript: N.G.S., E.Ş., B.Ç.U.; Critical Revision of Manuscript: N.G.S., E.Ş., B.Ç.U.; Final Approval and Accountability: N.G.S., E.Ş., B.Ç.U.

Conflict of Interest: The authors have no conflict of interest to declare.

Grant Support: The authors declared that this study has received no financial support.

References

Abdullateef, A. O., & Salleh, S. M. (2013). Does customer relationship management influence call center quality performance? An empirical industry analysis. Total Quality Management & Business Excellence, 24(9-10), 1035-1045.

Afrapoli, A. M., Tabesh, M., & Askari-Nasab, H. (2019). A multiple objective transportation problem approach to dynamic truck dispatching in surface mines. European Journal of Operational Research, 276(1), 331-342.

Aktekin, T. (2014). Call center service process analysis: Bayesian parametric and semi-parametric mixture modeling. European Journal of Operational Research, 234(3), 709-719.

Alotaibi, Y., & Liu, F. (2013). Average waiting time of customers in a new queue system with different classes. Business Process Management Journal.

Amorim-Lopes, M., Guimarães, L., Alves, J., & Almada-Lobo, B. (2021). Improving picking performance at a large retailer warehouse by combining probabilistic simulation, optimization, and discrete-event simulation. International Transactions in Operational Research, 28(2), 687-715.

Andrade, C. (2019). The P-value and statistical significance: misunderstandings, explanations, challenges, and alternatives. Indian Journal of Psychological Medicine, 41(3), 210-215.

Anton, J. (2000). The past, present, and future of customer access centers. International Journal of Service Industry Management.

Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data-enabled: Personalization & customization of services. Applied Computing and Informatics, 15(2), 94-101.

Anuşlu, M. D., & Fırat, S. Ü. (2019). Clustering analysis application on Industry 4.0-driven global indexes. Procedia Computer Science, 158, 145-152.

Arora, A. (2007). Dynamic project management using simulations. Project Managment Institute.

Avramidis, A. N., & L'Ecuyer, P. (2005, December). Modeling and simulation of call centers. In Proceedings of the Winter Simulation Conference, 2005. (pp. 9-pp). IEEE.

Banks, J., Carson II, J. S., & Barry, L. (2005). Discrete-event system simulation fourth edition

Behiri, W., Belmokhtar-Berraf, S., & Chu, C. (2018). Urban freight transport using passenger rail network:

Scientific issues and quantitative analysis. Transportation Research Part E: Logistics and Transportation Review, 115, 227-245.

Bholowalia, P., & Kumar, A. (2014). EBK-means: A clustering technique based on the Elbow method and k-means in WSN. International Journal of Computer Applications, 105(9).

Bongomin, O., Mwasiagi, J. I., Nganyi, E. O., & Nibikora, I. (2020). A complex garment assembly line balancing using simulation-based optimization. Engineering Reports, 2(11), e12258.

Calis, B. (2016). Agent-Based Simulation Model for Profit Maximization. *Journal of Management and Information Science*, *4*(1), 26-33.

Carnein, M., & Trautmann, H. (2019, April). Customer segmentation based on transactional data using stream clustering. In Pacific-Asia Conference on Knowledge Discovery and Data Mining (pp. 280-292). Springer, Cham.

Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. Accounting, Auditing and Finance, 1(1), 5-8.

David, F. R. (2013). Strategic Management: Concepts and Cases. Pearson.

Dillman, D. A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J., & Messer, B. L. (2009). Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the Internet. Social science research, 38(1), 1-18.

Doomun, R., & Jungum, N. V. (2008). Business process modelling, simulation and reengineering: call centres. Business Process Management Journal.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of Business Research, 69(2), 897-904.

Farajian, M. A., & Mohammadi, S. (2010). Mining the banking customer behavior using clustering and association rules methods.

Farruh, K. (2019). Consumer Life Cycle and Profiling: A Data Mining Perspective. In Consumer Behavior and Marketing. IntechOpen.

Feinberg, R., De Ruyter, K., & Bennington, L. (2005). Cases in call center management: great ideas (th) at work. Purdue University Press.

Feinberg, R. A., Kim, I. S., Hokama, L., De Ruyter, K., & Keen, C. (2000). Operational determinants of caller satisfaction in the call center. International Journal of Service Industry Management.

Figueiredo, V., Duarte, F. J., Rodrigues, F., Vale, Z., & Gouveia, J. (2003, September). Electric energy customer characterization by clustering. In Proc. ISAP.

Gayathri, M., Jha, S., Parmar, M., & Malathy, C. (2020, February). Customer Profiling Using Demographic Analysis by Video Face Detection and Recognition. In 2020 International Conference on Inventive Computation Technologies (ICICT) (pp. 570-575). IEEE.

Gittins, P., McElwee, G., & Tipi, N. (2020). Discrete event simulation in livestock management. Journal of Rural Studies, 78, 387-398.

Greco, F., & Polli, A. (2020). Emotional Text Mining: Customer profiling in brand management. International Journal of Information Management, 51, 101934.

Gustriansyah, R., Suhandi, N., & Antony, F. (2020). Clustering optimization in RFM analysis based on k-means. Indonesia. J. Electr. Eng. Comput. Sci, 18(1), 470-477.

Hahnke, J. (2000). The CRM Lifecycle–Without CRM Analytics, Your Customers Won’t Even Know You’re There. Defying the Limits, 159-164.

Hassan, M. M. T. M., & Tabasum, M. (2018). Customer profiling and segmentation in retail banks using data mining techniques. International journal of advanced research in computer science, 9(4).

Hung, P. D., Lien, N. T. T., & Ngoc, N. D. (2019, March). Customer segmentation using hierarchical agglomerative clustering. In Proceedings of the 2019 2nd International Conference on Information Science and Systems (pp. 33-37).

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

Jintana, J., & Mori, T. (2019). Customer clustering for a new method of marketing strategy support within the courier business. Academia Book Chapter, 31(2), 1-19.

Kadir, M. A., & Achyar, A. (2019). Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku. id.

Klement, P., & Snášel, V. (2011). Using SOM in the performance monitoring of the emergency call-taking system. Simulation Modelling Practice and Theory, 19(1), 98-109.

Knapčíková, L., Behúnová, A., & Behún, M. (2020). Using a discrete event simulation as an effective method applied in the production of recycled material. Advances in Production Engineering & Management, 15(4), 431-440.

Koole, G., & Pot, A. (2006). An overview of routing and staffing algorithms in multi-skill customer contact centers.

Lam, K., & Lau, R. S. M. (2004). A simulation approach to restructuring call centers. Business Process Management Journal.

Legros, B., Jouini, O., & Koole, G. (2017). A uniformization approach for the dynamic control of queueing systems with abandonments. Operations Research, 66(1), 200-209

Liu, F., & Deng, Y. (2020). Determine the number of unknown targets in Open World based on Elbow method. IEEE Transactions on Fuzzy Systems.

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

Maheshwari, K., Khapekar, R., Bahl, A., & Bhatia, K. (2019). Credit Profile of E-Commerce Customer.

Maraghi, M., Adibi, M. A., & Mehdizadeh, E. (2020). Using RFM Model and Market Basket Analysis for Segmenting Customers and Assigning Marketing Strategies to Resulted Segments. Journal of Applied Intelligent Systems and Information Sciences, 1(1), 35-43.

Mehrotra, V., & Fama, J. (2003, December). Call center simulation modeling: methods, challenges, and opportunities. In Proceedings of the 35th conference on Winter simulation: driving innovation (pp. 135-143). Winter Simulation Conference.

Monks, T., Currie, C. S., Onggo, B. S., Robinson, S., Kunc, M., & Taylor, S. J. (2019). Strengthening the reporting of empirical simulation studies: Introducing the STRESS guidelines. Journal of Simulation, 13(1), 55-67.

Mousavi, S., Boroujeni, F. Z., & Aryanmehr, S. (2020). Improving customer clustering by optimal selection of cluster centroids in k-means and k-medoids algorithms. Journal of Theoretical and Applied Information Technology, 98(18).

Nainggolan, R., Perangin-angin, R., Simarmata, E., & Tarigan, A. F. (2019, November). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. In Journal of Physics: Conference Series (Vol. 1361, No. 1, p. 012015). IOP Publishing.

Namvar, M., Gholamian, M. R., & KhakAbi, S. (2010, January). A two phase clustering method for intelligent customer segmentation. In 2010 International Conference on Intelligent Systems, Modelling and Simulation (pp. 215-219). IEEE.

Niyagas, W., Srivihok, A., & Kitisin, S. (2006). Clustering e-banking customers using data mining and marketing segmentation. ECTI Transactions on Computer and Information Technology (ECTI-CIT), 2(1), 63-69.

Nugraha, J. A. M. (2020). Application of K-Means Algorithm for Customer Grouping. International Journal of Computer Theory and Engineering, 12(2).

Nwogu, E. C., Iwueze, I. S., & Nlebedim, V. U. (2016). Some tests for seasonality in time series data. Journal of Modern Applied Statistical Methods, 15(2), 24.

Rajagopal, D. (2011). Customer data clustering using data mining technique. arXiv preprint arXiv:1112.2663.

Riskadayanti, O., & Hisjam, M. (2019, April). Discrete-event simulation of a production process for increasing the efficiency of a newspaper production. In IOP Conference Series: Materials Science and Engineering (Vol. 495, No. 1, p. 012026). IOP Publishing.

Robinson, G., & Morley, C. (2006). Call centre management: responsibilities and performance. International Journal of Service Industry Management.

Rojlertjanya, P. (2019). Customer Segmentation Based on the RFM Analysis Model Using K-Means Clustering Technique: A Case of I.T. Solution and Service Provider in Thailand.

Rudskaia, E., & Eremenko, I. (2019). Digital clustering in customer relationship management. In E3S Web of Conferences (Vol. 135, p. 04010). EDP Sciences.

Rygielski, C., Wang, J. C., & Yen, D. C. (2002). Data mining techniques for customer relationship management. Technology in society, 24(4), 483-502.

Saberi, M., Hussain, O. K., & Chang, E. (2017). Past, present and future of contact centers: a literature review. Business Process Management Journal.

Sabuncu, İ., Türkan, E., & Polat, H. (2020). Customer Segmentation And Profiling With RFM Analysis. Turkish Journal of Marketing, 5(1), 22-36.

Sağlam, B., Salman, F. S., Sayın, S., & Türkay, M. (2006). A mixed-integer programming approach to the clustering problem with an application in customer segmentation. European Journal of Operational Research, 173(3), 866-879.

Sargent, R. G. (2013). Verification and validation of simulation models. Journal of simulation, 7(1), 12-24.

Shih, M. Y., Jheng, J. W., & Lai, L. F. (2010). A two-step method for clustering mixed categorical and numeric data. Tamkang Journal of Science and Engineering, 13(1), 11-19.

Shih, Y. Y., & Liu, C. Y. (2003). A method for customer lifetime value ranking—Combining the analytic hierarchy process and clustering analysis. Journal of Database Marketing & Customer Strategy Management, 11(2), 159-172.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018, April). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In IOP Conference Series: Materials Science and Engineering (Vol. 336, No. 1, p. 012017). IOP Publishing.

Thomas, M. R., & Shivani, M. P. (2020). Customer Profiling of Alpha. Ushus Journal of Business Management, 19(1), 75-86.

Troncoso-Palacio, A., Neira-Rodado, D., Ortíz-Barrios, M., Jiménez-Delgado, G., & Hernández-Palma, H. (2018, June). Using discrete-event-simulation for improving operational efficiency in laboratories: a case study in pharmaceutical industry. In International Conference on Swarm Intelligence (pp. 440-451). Springer, Cham.

Tsiptsis, K. K., & Chorianopoulos, A. (2011). Data mining techniques in CRM: inside customer segmentation. John Wiley & Sons

Umargono, E., Suseno, J. E., & Gunawan, S. V. (2020, October). K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula. In The 2nd International Seminar on Science and Technology (ISSTEC 2019) (pp. 121-129). Atlantis Press.

Uslu, B. Ç., & Fırat, S. Ü. O. (2019). A Comprehensive Study on Internet of Things Based on Key Artificial Intelligence Technologies and Industry 4.0. In *Advanced Metaheuristic Methods in Big Data Retrieval and Analytics* (pp. 1-26). IGI Global.

USLU, B. Ç. (2020). Capability model and competence measuring for smart hospital system: an analysis for turkey. *International Journal of Health Services Research and Policy*, *5*(1), 41-50.

Van Buuren, M., Kommer, G. J., van der Mei, R., & Bhulai, S. (2017). EMS call center models with and without function differentiation: A comparison. Operations Research for Health Care, 12, 16-28.

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., & Haenlein, M. (2019). Digital transformation: A multidisciplinary reflection and research agenda. Journal of Business Research.

Watson, J. (2012). The Requirements for Being an Analytics-Based Organization. Business Intelligence Journal, 17(2), 42-44.

Windarto, A. P., Siregar, M. N. H., Suharto, W., Fachri, B., Supriyatna, A., Carolina, I., ... & Toresa, D. (2019, August). Analysis of the K-Means Algorithm on Clean Water Customers Based on the Province. In Journal of Physics: Conference Series (Vol. 1255, No. 1, p. 012001). IOP Publishing.

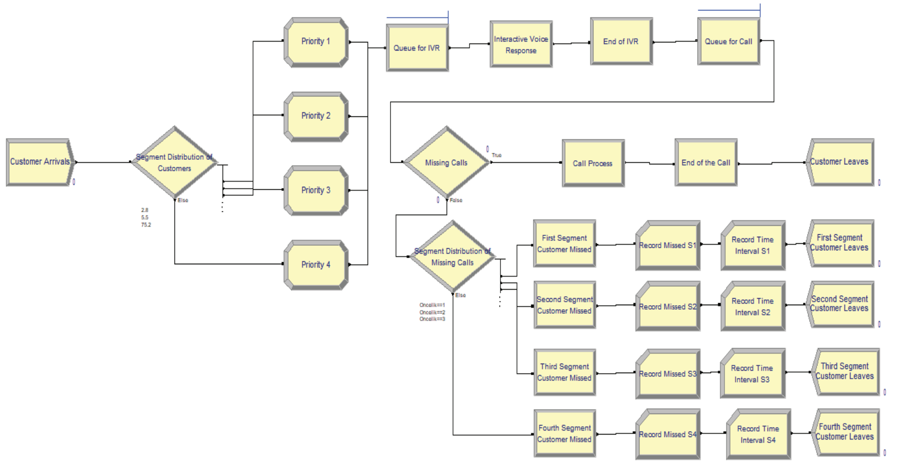
Windisch, J., Väätäinen, K., Anttila, P., Nivala, M., Laitila, J., Asikainen, A., & Sikanen, L. (2015). Discrete-event simulation of an information-based raw material allocation process for increasing the efficiency of an energy wood supply chain. Applied energy, 149, 315-325.

Ye, L., Qiuru, C., Haixu, X., Yijun, L., & Guangping, Z. (2013). Customer segmentation for telecom with the k-means clustering method. Information Technology Journal, 12(3), 409413.

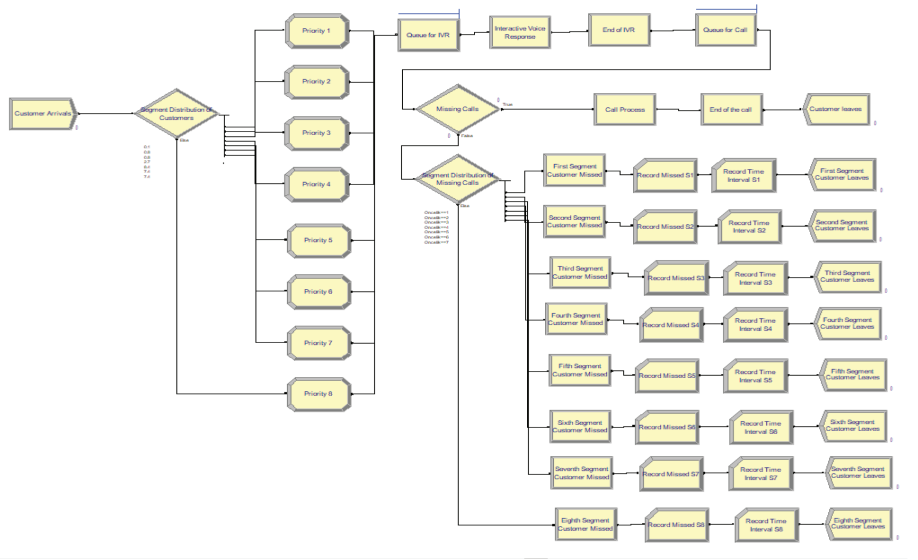
Yemane, A., Gebremicheal, G., Meraha, T., & Hailemicheal, M. (2020). Productivity improvement through line balancing by using simulation modeling. Journal of Optimization in Industrial Engineering, 13(1), 153-165.

Zalaghi, Z., & Varzi, Y. (2014). Measuring customer loyalty using an extended RFM and clustering technique. Management Science Letters, 4(5), 905-912.

APPENDIX



Appendix 1: Simulation model prepared according to company segments



Appendix 2: Simulation model prepared according to data analytics segments