

Aplicação de teoria das filas no atendimento telefônico: *configuração de turnos em uma autarquia federal*

Felipe Simoyama

Doutorando em Pesquisa Operacional pela Unifesp
Mestrado em Sistemas Complexos pela USP
Graduado em Administração pela UFSJ
E-mail: simoyama@unifesp.br

Daniela Martins Diniz

Doutora em Administração pela UFMG.
Professora Adjunta da UFSJ.
E-mail: danidiniz@ufsj.edu.br

Mario Cesar Guimarães Battisti

Graduação em Terapia Ocupacional pela PUC-Camp
Mestrado em Filosofia-Ética pela PUC-Camp
Gerente Administrativo no CREFITO-3
E-mail: battisti@uol.com.br

Recebido: 13 dez. 2021

Aprovado: 26 ago. 2022

Resumo: O presente estudo de caso teve o objetivo de aplicar conceitos de teoria das filas para analisar e propor soluções para otimização do atendimento telefônico em uma autarquia federal. Os resultados indicam a possibilidade de utilização de uma configuração de turno em que o tempo de espera do usuário seja, em média, de 21 segundos para ser atendido, proporcionando um tempo total de permanência do usuário no sistema (tempo de fila e atendimento) de 3,37 minutos.

Palavras-chave: Teoria das Filas. Estudo de Caso. Atendimento Telefônico.

Abstract: This case study aimed at applying queuing theory concepts to analyze and propose solutions for optimizing a call center in a Brazilian federal agency. Results show the possibility of using a shift configuration in which the user's waiting time is of 21 seconds to be serviced (on average), and the total time a user remains in the system (queue plus service time) is of 3.37 minutes.

Palavras-chave: Queuing Theory. Case Study. Call Center.

Resumen: El presente estudio de caso tuvo como objetivo aplicar conceptos de la teoría de colas para analizar y proponer soluciones para optimizar el servicio telefónico en una dependencia federal. Los resultados indican la posibilidad de utilizar una configuración de turnos en la que el tiempo de espera del usuario sea, en promedio, de 21 segundos para ser atendido, proporcionando un tiempo total de permanencia del usuario en el sistema (tiempo de cola y atención) de 3,37 minutos.

Palabras clave: Teoría de Colas. Estudio de Caso. Atendimento Telefonico.

1 Introdução

Os setores de atendimento são o canal de entrada para a prestação do serviço, tanto em instituições públicas quanto privadas. No setor público brasileiro, no entanto, trata-se de uma área cada vez mais crítica, por conta das limitações inerentes ao serviço público, como as pressões orçamentárias e o crescimento populacional, que levam ao aumento da demanda. Ainda, uma organização governamental, via de regra, deve prestar um atendimento universal (RUTKOWSKI, 1998), garantindo o acesso quase irrestrito aos seus serviços.

Os canais de atendimento ao público são variados. Atualmente, as instituições oferecem atendimentos presenciais, atendimentos via telefone/call center e diversas modalidades de atendimento eletrônico via internet, como o chat e aplicativos de mensagens instantâneas (ARAUJO, REINHARD, CUNHA, 2018, LEITE, 2017). Nesse sentido, o atendimento telefônico é uma forma de contato rápida e eficiente nessas instituições (MACCARINI, 2017), uma vez que ele elimina as barreiras de distância entre os interlocutores, ao passo que economiza tempo e recursos do cidadão e da própria instituição. O cidadão, ao tentar acessar um serviço público por meio do telefone, o mesmo espera um atendimento eficaz e, principalmente, ágil e objetivo. No entanto, tendo em vista que os órgãos públicos geralmente trabalham com recursos limitados, tanto em termos de pessoal quanto de infraestrutura, prestar um atendimento que atenda às expectativas do cidadão pode ser um grande desafio.

Para que o atendimento telefônico possa ser otimizado com tais recursos limitados, muitas organizações buscam ferramentas e tecnologias que permitam agilizar esse atendimento sem a necessidade de grandes investimentos ou contratações. Um desses exemplos é a utilização de simulações computacionais que permitam uma melhor configuração dos setores de atendimento, permitindo que um maior número de atendimentos possa ser realizado com a mesma quantidade de recursos (CAVALHEIRO, VAROLLO, BERTOLUCI, 2018).

Uma das possibilidades de se proceder com tal otimização é o uso de modelagem matemática baseada em teoria das filas combinada com ferramentas de simulação computacional (HOU, CAO, FAN, 2021), permitindo que o gestor analise diversos

cenários, tanto em termos de quantidade de atendimentos quanto em termos de diferentes configurações possíveis, para que possa tomar uma decisão bem fundamentada.

As filas fazem parte do cotidiano da maioria das pessoas. Sejam elas filas físicas (atendimento presencial), sejam elas filas virtuais, como é o caso de atendimentos telefônicos, chats e outros. Na pesquisa operacional, o conhecimento de teoria das filas é cada vez mais utilizado com objetivo de otimizar a utilização de recursos para a melhoria da prestação de serviços.

Este trabalho tem como objetivo apresentar um estudo de caso de modelagem baseada em teoria das filas e simulação computacional para otimização do atendimento telefônico de uma organização pública, bem como utilização de técnicas de pesquisa operacional (PO) *soft* na fase de estruturação do problema.

2 Referencial teórico

Os sistemas baseados em teoria das filas geralmente são descritos utilizando a notação de Kendall-Lee, desenvolvida inicialmente por Kendall (1953) e posteriormente complementada por Lee (1966). Nesse padrão, utilizam-se seis siglas que resumem o problema sob estudo (KIM, 2012, BAKARI, CHAMALWA, BABA 2014), conforme segue:

$$a/b/c : d/e/f \quad (1)$$

em que as siglas a, b, c, d, e, f correspondem aos elementos básicos de um modelo de fila, conforme segue:

- a = distribuição de probabilidade - processo de chegada (e.g. M = Poisson)
- b = distribuição de probabilidade - serviço realizado (e.g. M = exponencial)
- c = número de servidores em paralelo
- d = disciplina da fila (e.g. FCFS)
- e = número máximo de usuários no sistema (capacidade)
- f = tamanho da população (fonte de usuários)

A título de exemplo, um modelo com a notação $M/G/1 \{FCFS/\infty/\infty\}$, segundo a notação de Kendall-Lee, pode ser identificado como um modelo de filas que possui processo de chegada markoviano, com distribuição de probabilidade genérica no tempo de serviço, tendo apenas 1 servidor em paralelo, no modelo de First-Come First-Serve, em que há capacidade ilimitada de usuários no sistema e população infinita. Os leitores interessados em notação de teoria das filas podem encontrar informações em Cooper (2005).

3 Método

3.1 Estudos de caso em pesquisa operacional

O emprego de estudos de caso é corriqueiro no âmbito da pesquisa operacional. Há diversos exemplos de pesquisas na área que se valem da modelagem de problemas da vida real, geralmente estudando situações bastante específicas, para se realizar um processo de simulação ou otimização em um contexto aplicado.

Como exemplo, Marins, Souza e Barros (2009) realizaram um estudo de caso de aplicação do método de análise hierárquica em uma Prefeitura. Mais recentemente, Santos (2021) realizou um estudo de caso para minimização de custos em uma barbearia. No que se refere à teoria das filas, Bruns, Soncim e Sinay (2001) realizaram um estudo de caso para gestão de filas em uma loja de departamentos, enquanto que Morabito e Lima (2000) a utilizaram para um estudo sobre o problema de filas em caixas de supermercado.

Nota-se um grande número de publicações utilizando estudos de caso com aplicação de técnicas de pesquisa operacional em diversos setores diferentes, o que denota a relevância que a pesquisa operacional vem ganhando em processos de tomada de decisão mais racionais e com foco na eficiência da gestão.

3.2 Po soft para estruturação de problemas

Embora a PO tenha se originado e disseminado em abordagens hard, com enfoque em modelagens matemáticas e realização de simulações e otimizações pela via computacional, a abordagem soft tem sido cada vez mais utilizada na pesquisa operacional, especialmente no processo de estruturação dos problemas e, via de regra, em

questões de nível estratégico, onde os cenários são mais complexos e há maior subjetividade (ENSSLIN, VIANNA, 2008).

A PO hard é um método bastante adequado para lidar com problemas bem definidos e estruturados, ou que possam ser modelados de forma mais objetiva, ou mesmo determinística. No entanto, esse método possui limitações no que se refere a problemas que envolvem pessoas e maior complexidade especialmente em situações onde há visões diferentes/antagônicas sobre o problema (LIMA, MOLINARO, 2016). Para Pamplona e Lins (2020), já está consolidado na literatura o fato de que a PO hard e a PO soft não se contrapõem, mas são métodos complementares e ambos são necessários para se enfrentar os problemas de tomada de decisão da vida real.

Dentre os métodos mais utilizados para estruturação de problemas está o SSM (Soft Systems Methodology), desenvolvido por Checkland e que talvez seja o método mais utilizado para estruturação de problemas em PO Soft (HEYER, 2004). O SSM é um método de modelagem de problemas aplicável a diversas áreas. O método passou por algumas revisões ao longo do tempo, incluindo uma simplificação feita pelo próprio Checkland (YAN, 2009), bem como adaptações que podem ser vistas em outros autores, como Sankaran et al. (2009).

O presente estudo apresenta uma abordagem simplificada do SSM semelhante àquela encontrada em Williams e Gunatunge (1999). Aqui, o método é utilizado para uma melhor compreensão do sistema sob estudo, suas limitações e restrições, bem como uma análise sobre o envolvimento das partes interessadas nesse sistema.

4 Resultados

4.1 Caso estudado

A organização sob estudo se trata de uma autarquia federal, mais especificamente um conselho de fiscalização profissional, localizado na cidade de São Paulo/SP e que possui 10 subsedes (filiais) localizadas no interior do Estado de São Paulo. Essa autarquia é responsável pelo registro e fiscalização de mais de 80.000 profissionais (pessoas físicas) da área da saúde, além de mais de 10.000 pessoas jurídicas. Os departamentos de registro e fiscalização, bem como do departamento financeiro da autarquia, são aqueles que mais geram demandas de serviço e atendimento, especialmente os contatos por telefone.

O setor de atendimento telefônico da instituição conta com 7 (sete) funcionários, possuindo uma demanda de mais de 60.000 atendimentos anuais. Tais ligações têm uma média de 2,84 minutos de duração, porém, com um desvio padrão de 2,27 minutos. Dentre os principais indicadores de desempenho do setor, está o Índice de Chamadas Perdidas, que representa quantas ligações dos usuários sequer chegam a ser atendidas por um funcionário. Esse índice é afetado tanto pela sazonalidade das chamadas nos meses iniciais do ano, como por eventos esporádicos que podem gerar uma demanda acima do normal, sendo exemplificado na figura pelo mês de maio (14,93%). Esses eventos podem ser, por exemplo, um comunicado enviado em lote aos jurisdicionados.

Por ser considerado o principal canal de entrada das demandas da organização, há uma preocupação constante da autarquia com relação à satisfação deste serviço. Nesse sentido, é realizada uma pesquisa de satisfação com o atendimento telefônico rotineiramente, sendo medida através de contato ativo feito por um colaborador da autarquia, o qual realiza ligações para uma amostra de cerca de 2% (dois por cento) de todas as ligações que foram atendidas. Embora os índices de satisfação estejam invariavelmente altos (sempre acima de 94%), a satisfação é ligeiramente menor em meses como o de maio, quando o índice de chamadas perdidas também é alto.

O horário de atendimento telefônico de que trata o presente estudo ocorre entre 8h30 e 16h30, contando com uma equipe de 7 (sete) funcionários que se revezam em turnos de 6h cada, conforme apresentado no Apêndice 1. Cada funcionário (telefonista) tem uma carga horária de 6h diárias de trabalho. Como a organização sob estudo permite que as jornadas dos telefonistas sejam iniciadas dentro de 3 possibilidades, conclui-se que há 3 (três) diferentes turnos possíveis nos quais os telefonistas podem ser alocados, a saber:

- T1: 8h30 às 14h30
- T2: 9h30 às 15h30
- T3: 10h30 às 16h30

A partir do Apêndice 1, pode-se afirmar que 2 funcionários estão alocados em T1, 2 funcionários alocados em T2 e 3 funcionários alocados em T3. Essa configuração pode ser denominada por 2, 2, 3. O Quadro 1 visa tornar claro as possibilidades de alocação.

Quadro 1: Turnos possíveis para os servidores.

| Turno | 8:30 | 9:30 | 10:30 | 11:30 | 12:30 | 13:30 | 14:30 | 15:30 |
|-------|------|-------|-------|-------|-------|-------|-------|-------|
| | 9:30 | 10:30 | 11:30 | 12:30 | 13:30 | 14:30 | 15:30 | 16:30 |
| T1 | | | | | | | | |
| T2 | | | | | | | | |
| T3 | | | | | | | | |

Utilizando análise combinatória, é possível verificar que o número de permutações possíveis de 8 elementos combinados 3 a 3 gera um total de 512 combinações possíveis. Importante notar que, para esse cálculo, foram considerados 8 elementos, uma vez que cada turno pode ter de 1 a 7 funcionários, mas também existe a possibilidade de não alocar ninguém em um determinado turno. No entanto, esse tipo de análise resulta em alocações não desejáveis, uma vez que permite a formação de equipes sem a utilização de todos os funcionários (ociosidade).

Nesse caso, é preciso colocar uma restrição na análise, de forma que as equipes e turnos formados sempre considerem a utilização de toda a equipe (7 funcionários no caso). Para isso, foi desenvolvido um script simples em Python para que não apenas mostrasse a quantidade de combinações possíveis, mas também como são formadas tais alocações. O resultado do script é uma lista de turnos possíveis. Essa lista é apresentada no Apêndice 2.

A tabela acima aponta que há 36 possibilidades de alocação da equipe de 7 funcionários em 3 turnos diferentes. A nomenclatura que será utilizada daqui em diante indica, por exemplo, que a alocação (2,4,1) corresponde a 2 funcionários no turno que se inicia às 8h30, 4 funcionários no turno que se inicia às 9h30 e 1 funcionário no turno que se inicia às 10h30.

No entanto, alguns turnos são considerados inviáveis, pois se fossem implementados, deixariam uma ou mais faixas de horário descobertas (Tabela). Uma das restrições do presente estudo de caso é a manutenção do horário de funcionamento do Pronto Atendimento. Dessa forma, torna-se inviável a alocação de turnos sem cobertura de atendentes.

4.2 Splicação do SSM

O objetivo desta seção é, primeiramente, apresentar os resultados da aplicação do framework SSM simplificado, onde foi feita a estruturação do problema. Na segunda parte desta seção, serão apresentados os resultados da modelagem matemática do problema, utilizando teoria das filas, bem como o resultado do processo de otimização.

A Figura Rica, que é desenvolvida nas duas primeiras etapas do SSM, é um desenho, geralmente feito manualmente, onde os participantes de uma reunião analisam o problema sob estudo a partir de diferentes visões das partes interessadas. Na primeira etapa, é realizado um esboço da referida figura, de forma a compreender o envolvimento das partes interessadas no sistema em estudo. Em seguida, são feitas análises e interpretações sobre o sistema, para que se possa chegar a uma Figura Rica em sua versão final.

O Quadro 2, por sua vez, apresenta o resultado da etapa 3 do SSM, onde é definida a tabela CATWOE.

Quadro 2: Definição CATWOE do sistema.

| Sigla | Definição |
|-------|---|
| C | O cliente (Customer) é a sociedade jurisdicionada |
| A | Os atores (Actors) são os atendentes e aqueles que efetivamente utilizam o sistema |
| T | O processo de transformação (Transformation) é o atendimento célere aos usuários |
| W | A visão de mundo (Worldview) é a necessidade de realizar o processo de atendimento dentro de prazo razoável e que não gere insatisfação |
| O | O dono (Owner) é a diretoria da autarquia e o chefe do departamento |
| E | As restrições ambientais (Environmental constraint) são a própria alta demanda de atendimentos combinada com a impossibilidade de aumento de recursos no curto prazo. |

A partir da figura rica combinada com a tabela CATWOE, são feitas as seguintes proposições em discussão com as partes envolvidas.

- O sistema de Pronto Atendimento da autarquia é o principal canal para recepção de demanda dos usuários, sendo responsável por cerca de 60 mil atendimentos por ano. Trata-se de um canal de atendimento receptivo, com ligações que geralmente não passam de 5 minutos. Quando as dúvidas não podem ser resolvidas de imediato para o usuário, é aberto um chamado para que o atendente de 2º nível dê uma resposta adequada em até 1 (um) dia útil.

Aplicação de teoria das filas no atendimento telefônico

- Embora os níveis de satisfação com o setor aparentem estar elevados, não há uma certeza ou segurança sobre o indicador, uma vez que há reclamações feitas em redes sociais e o método de apuração do indicador deixe de lado os clientes potencialmente mais insatisfeitos, ou seja, aqueles que sequer conseguiram ser atendidos (ver índice de ligações perdidas).
- De forma geral, a insatisfação é maior com relação ao prazo de atendimento. Os clientes querem ser atendidos com a maior celeridade possível. O tempo de atendimento é um fator crítico para o setor.
- Por sua vez, as tomadas de decisão devem ser planejadas e, por se tratar de administração pública, as decisões devem ser tomadas de forma transparente e da forma mais objetiva possível.

Assim, na etapa 4 são lançados os seguintes desafios:

- Compreender melhor o modelo de trabalho atual.
- Analisar quais são os recursos disponíveis atualmente.
- Levantar dados com a maior completude possível dos atendimentos já realizados, especialmente em relação ao tempo.
- Discutir quais são as possibilidades de alcançar os objetivos propostos considerando a impossibilidade de investimentos no curto prazo.

Nas etapas 5 e 6, o modelo desenvolvido é confrontado com o mundo real, onde, para cada proposição acima, são feitas reflexões para aprofundamento. Nessa etapa, foi definido que, para a primeira e segunda proposições (modelo atual), seriam realizadas reuniões com o chefe do Pronto Atendimento, de forma a aprofundar o conhecimento sobre o problema enfrentado pela organização (primeira proposição) e sobre a análise dos recursos disponíveis (segunda proposição).

O levantamento de dados foi feito por meio de extração do software de gerenciamento de call center disponível na autarquia. Nessa etapa, foi observado que há dados disponíveis de todas as ligações recebidas pelo Pronto Atendimento a partir de agosto de 2019 até os dias atuais. No entanto, os dados de 2020 apresentam dados do período em que os servidores foram colocados sob regime de teletrabalho provisoriamente, motivo pelo qual optou-se pela utilização de dados apenas entre agosto e dezembro de 2019, uma vez que foi observado se tratar de uma quantidade bastante significativa (mais de 20 mil ligações no período).

Já com relação à quarta proposição (discussão das possibilidades), foi constatado que, diante da impossibilidade de realizar investimentos e contratações no curto prazo, a melhor alternativa seria analisar a possibilidade de otimização do setor utilizando os recursos atualmente disponíveis, sendo possível a alteração de escalas e turnos dos servidores. Assim, optou-se pela análise dos dados utilizando conhecimentos de teoria das filas e modelo de configuração de turnos.

Por fim, a etapa 7 do SSM corresponde à tomada de ação para colocar as propostas de mudança em prática, o que será discutido nas próximas seções deste estudo.

4.3 Modelagem matemática e solução do problema

O problema foi modelado pela notação de Kendall-Lee para o padrão $M/M/m\{FCFS/\infty/\infty\}$, com uma fila única e k servidores. Os parâmetros para calcular o desempenho do sistema são apresentados no Quadro 3:

Quadro 3: parâmetros para análise de desempenho.

| Sigla | Significado | Equação |
|--------|---|--|
| r | Taxa de ocupação do servidor | |
| ρ | Taxa de Ocupação do sistema | |
| NS | Número médio de usuários no sistema | |
| TS | Tempo médio de permanência no Sistema | |
| NF | Número médio de clientes na Fila (aguardando) | $\frac{(\rho m)^m \rho}{m!(1-\rho)^2} P_0$ |
| TF | Tempo médio do cliente na Fila | |
| | Ociosidade total | $\frac{1}{\sum_{n=0}^{m-1} \frac{(\rho m)^n}{n!} + \frac{(\rho m)^m}{m!} \left(\frac{1}{1-\rho} \right)}$ |

Para calcular o desempenho do sistema e otimizar os turnos, primeiramente é necessário calcular os valores de λ (intervalo médio entre chegadas) e μ (taxa de serviço) para intervalos horários, cujos valores estão apresentados no Quadro 4.

Quadro 4: intervalo médio entre chegadas (λ) e tempo de serviço (μ) a cada intervalo de 1h.

| Turno | 8h30-9h30 | 9h30-10h30 | 10h30-11h30 | 11h30-12h30 | 12h30-13h30 | 13h30-14h30 | 14h30-15h30 | 15h30-16h30 |
|-----------|-----------|------------|-------------|-------------|-------------|-------------|-------------|-------------|
| λ | 2,04 | 1,94 | 1,92 | 2,00 | 2,35 | 2,27 | 2,25 | 2,33 |
| μ | 3,16 | 2,88 | 2,81 | 2,80 | 2,87 | 2,87 | 3,27 | 3,45 |

A título de exemplo, a tabela acima indica que o período em que há menos ligações é entre 12h30-13h30, quando ocorre 1 ligação a cada 2,35 minutos (2m21s). Nesse mesmo período, as ligações duram, em média, 2,87 minutos (2m52s).

Com esses valores, é possível calcular o desempenho do sistema para cada intervalo de 1 hora para, em seguida, analisar qual o turno que retorna o melhor resultado. Para fins desse estudo, será considerada a menor taxa de ocupação média para cada configuração de turno, dentre as 36 possibilidades existentes, conforme Apêndice 2.

A partir do Quadro 4, inicialmente, foram calculadas as taxas de ocupação ρ para cada faixa de horário com 1 a m atendentes. Esses resultados são apresentados no Apêndice 3. Em todas as faixas de horário, verifica-se a inviabilidade de se alocar apenas 1 funcionário ($m=1$), pois em todos esses casos observa-se $\rho > 1$, tornando o sistema instável, o que levaria a fila a crescer indefinidamente. O sistema é considerado instável quando a condição $\rho < 1$ não é atendida. Essa é uma condição necessária para sistemas sem limite de entrada de usuários, como o do presente estudo.

Em modelos do tipo M/M/1/K, quando existe uma limitação de K usuários presentes no sistema, é possível representar o sistema por um modelo conhecido como “nascimento e morte”, cujos valores de P podem ser representados pelas fórmulas abaixo e a condição $\rho < 1$ não é mais necessária para que essa série convirja (ARENALES et al., 2015). Porém, no estudo de caso tal limitação não existe, pois o modelo do tipo M/M/m{/FCFS/ ∞/∞ } permite número infinito de chegada de usuários a partir de uma população também infinita.

Com o objetivo de facilitar a compreensão da relação entre turnos e horários aceitáveis, foi elaborado o Apêndice 4. A coluna “Turno” indica o número de funcionários que seriam alocados em cada turno. As colunas dos horários apontam quantos funcionários estariam em serviço nos respectivos turnos. A coluna aceitável significa “Não” quando pelo menos um horário estiver descoberto (sem nenhum funcionário). A coluna estável indica se em todos os horários a condição $\rho < 1$ foi atendida. Por fim, a coluna viável indica quais turnos atendem às duas condições (aceitabilidade e estabilidade).

O Apêndice 4 indica, portanto, que 10 (dez) possibilidades de turnos passaram pelas restrições impostas. Esses turnos foram analisados para o valor médio de ρ das 8 faixas de horário que compõem o expediente. O Quadro 5 apresenta esses resultados.

Quadro 5: análise de desempenho em função de ρ para os turnos viáveis.

| Índice | Turno | 8:30-9:30 | 9:30-10:30 | 10:30-11:30 | 11:30-12:30 | 12:30-13:30 | 13:30-14:30 | 14:30-15:30 | 15:30-16:30 | Valor médio de ρ |
|--------|-----------|-----------|------------|-------------|-------------|-------------|-------------|-------------|-------------|-----------------------|
| 16 | (2, 0, 5) | 2 | 2 | 7 | 7 | 7 | 7 | 5 | 5 | 0,358468 |
| 17 | (2, 1, 4) | 2 | 3 | 7 | 7 | 7 | 7 | 5 | 4 | 0,336794 |
| 18 | (2, 2, 3) | 2 | 4 | 7 | 7 | 7 | 7 | 5 | 3 | 0,336754 |
| 19 | (2, 3, 2) | 2 | 5 | 7 | 7 | 7 | 7 | 5 | 2 | 0,358324 |
| 22 | (3, 0, 4) | 3 | 3 | 7 | 7 | 7 | 7 | 4 | 4 | 0,313607 |
| 23 | (3, 1, 3) | 3 | 4 | 7 | 7 | 7 | 7 | 4 | 3 | 0,313566 |
| 24 | (3, 2, 2) | 3 | 5 | 7 | 7 | 7 | 7 | 4 | 2 | 0,335136 |
| 27 | (4, 0, 3) | 4 | 4 | 7 | 7 | 7 | 7 | 3 | 3 | 0,312570 |
| 28 | (4, 1, 2) | 4 | 5 | 7 | 7 | 7 | 7 | 3 | 2 | 0,334139 |
| 31 | (5, 0, 2) | 5 | 5 | 7 | 7 | 7 | 7 | 2 | 2 | 0,354735 |

Assim, a configuração de turno de número 27 (4, 0, 3) foi a que apresentou o melhor resultado no teste, apresentando a menor taxa de ocupação média ($\rho = 31,26\%$). Nessa configuração, sugere-se que 4 funcionários trabalhem no Turno 1 (período entre 8h30 e 14h30) e 3 funcionários trabalhem no Turno 3 (período entre 10h30 e 16h30). Ressalta-se que a opção 4, 0, 3 é uma configuração viável, mesmo com nenhum funcionário T2, pois não deixa nenhum horário descoberto de atendimento, o que é melhor ilustrado pelo Quadro 1. Tal configuração apresenta um resultado superior à configuração 18 (2, 2, 3), que era utilizado na autarquia à época, com uma taxa de ocupação superior ($\rho = 33,68\%$).

O Quadro 6 apresenta a média de todos os indicadores de desempenho do sistema para os turnos viáveis.

Quadro 6: indicadores de desempenho dos turnos viáveis

| Índice | Turno | ρ | NS | TS | NF | TF | |
|--------|-----------|--------|--------|--------|--------|--------|--------|
| 16 | (2, 0, 5) | 0,3585 | 2,7613 | 4,3353 | 0,6612 | 1,3215 | 0,7739 |
| 17 | (2, 1, 4) | 0,3368 | 2,4190 | 3,9243 | 0,4466 | 0,9106 | 0,7898 |
| 18 | (2, 2, 3) | 0,3368 | 2,4180 | 3,9412 | 0,4462 | 0,9274 | 0,7899 |
| 19 | (2, 3, 2) | 0,3583 | 2,7545 | 4,4392 | 0,6576 | 1,4254 | 0,7740 |
| 22 | (3, 0, 4) | 0,3136 | 1,9673 | 3,3680 | 0,1726 | 0,3543 | 0,8055 |
| 23 | (3, 1, 3) | 0,3136 | 1,9663 | 3,3849 | 0,1722 | 0,3711 | 0,8055 |
| 24 | (3, 2, 2) | 0,3351 | 2,3028 | 3,8829 | 0,3836 | 0,8691 | 0,7896 |
| 27 | (4, 0, 3) | 0,3126 | 1,9414 | 3,3727 | 0,1619 | 0,3590 | 0,8065 |
| 28 | (4, 1, 2) | 0,3341 | 2,2779 | 3,8707 | 0,3733 | 0,8570 | 0,7906 |
| 31 | (5, 0, 2) | 0,3547 | 2,5687 | 4,2988 | 0,5620 | 1,2851 | 0,7766 |

Conclusão

Este estudo apresentou uma aplicação de teoria das filas combinada com SSM no estudo de caso do setor de atendimento telefônico de um órgão da administração pública, em um cenário com limitações de recurso, tanto humanos quanto físicos, diante de não

haver previsão para novas contratações de servidores e, ainda que houvesse, não haveria espaço físico para alocá-los para uma eventual expansão do serviço. Essa classe de problemas é encontrada na literatura de pesquisa operacional como *shift planning*, *shift configuration*, *employee scheduling* ou *timetabling* (KABAK, 2008). No entanto, geralmente são encontrados problemas diferentes desse estudo de caso, com regime de escalas, amplamente utilizados na indústria e no setor de saúde.

A aplicação de conceitos de pesquisa operacional, tais como a teoria das filas e otimização, são de grande utilidade justamente nesse tipo de situação, onde há notória escassez de recursos, com diversos exemplos existentes na literatura, como pode ser visto, por exemplo, em Rauner et al (2003) e Carravilla e Oliveira (2013). As técnicas de pesquisa operacional soft, como o SSM, ainda que de forma simplificada, permitiram que a modelagem do problema de forma mais assertiva em relação ao que é praticado no órgão sob estudo, como a identificação das restrições citadas no parágrafo anterior, bem como a restrição da não possibilidade de mudança do horário de expediente atualmente praticado na autarquia e o melhor entendimento da lógica de horários e turnos então praticada.

A Figura Rica tornou mais clara a percepção de que a melhoria no tempo de atendimento do serviço poderia ser alcançada apenas por meio de uma mudança sem impactos orçamentários e/ou financeiros. Além disso, houve de se levar em conta a cobrança dos cidadãos por maior efetividade no serviço e a cobrança dos órgãos de controle externo para que as decisões sejam tomadas com o devido planejamento. Nesse sentido, a modelagem do problema foi feita de tal forma que fosse feita uma otimização do resultado tão somente a partir de uma mudança de processo, sem a necessidade de novos recursos de qualquer sorte (físicos, humanos, financeiros etc.), o que permite uma melhoria imediata no setor que é o principal canal de atendimento da autarquia.

Os resultados permitiram verificar qual é a configuração de turnos que melhor balanceia a relação entre a taxa de entrada dos clientes na fila do atendimento telefônico, a duração desses atendimentos e a disponibilidade dos atendentes. Esse balanceamento tende a diminuir a taxa de ocupação do sistema, proporcionando a redução nos tempos de espera, a maior resiliência do sistema para períodos de eventual demanda extraordinária e, conseqüentemente, maior satisfação do usuário. Como exemplo, a melhor configuração de turno resultante do modelo viabiliza um atendimento com tempo médio de espera de 0,359 minutos (21 segundos). Comparando-se esse resultado com a configuração de turnos então utilizada pela autarquia, à época da coleta de dados, a diferença é

significativa: 0,9274 minutos, ou 55 segundos - representando uma diminuição de 61,29% do tempo médio.

Uma das limitações desse estudo é o fato de utilizar dados apenas do ano de 2019. Não foi possível utilizar dados anteriores por conta da indisponibilidade de dados, uma vez que não havia sistema de gerenciamento (contact center) implantado no órgão sob estudo. Da mesma forma, não foi possível utilizar dados mais recentes, como do ano de 2020, uma vez que os servidores da autarquia passaram a atuar provisoriamente em regime de teletrabalho, o que impactou significativamente o setor e, conseqüentemente, os usuários. Além disso, embora a autarquia seja responsável pelo atendimento de mais de 80.000 jurisdicionados, verifica-se que, na prática, trata-se de uma instância pequena para problemas de pesquisa operacional mais modernos, sendo que o presente estudo de caso pôde ser solucionado com utilização de planilhas eletrônicas. Porém, problemas com instâncias maiores podem requerer a elaboração de um modelo matemático de otimização, a utilização de solvers modernos ou mesmo a elaboração de heurísticas para solucioná-lo.

Sugere-se, para estudos futuros, a utilização de teoria das filas combinada com técnicas de otimização para solucionar problemas de configuração de turnos com instâncias maiores e possibilidades de aplicação para outros setores que careçam de tais ferramentas.

Apêndices

Os apêndices deste artigo estão disponíveis em: <https://doi.org/10.5281/zenodo.7098213>.

Referências

ARAÚJO, M. H. de; REINHARD, N.; CUNHA, M. A. Serviços de governo eletrônico no Brasil: uma análise a partir das medidas de acesso e competências de uso da internet. **Revista de Administração Pública**, v. 52, n. 4, p. 676-694, 2018.

ARENALES, M., ARMENTANO, V., MORABITO, R., & YANASSE, H. **Pesquisa operacional: para cursos de engenharia**. São Paulo. Elsevier Brasil. 2015

BAKARI, H. R.; CHAMALWA, H. A.; BABA, A. M. Queuing process and its application to customer service delivery (A case study of Fidelity Bank Plc, Maiduguri). **International Journal of Mathematics and Statistics Invention (IJMSI)**, v. 2, n. 1, p. 14-21, 2014.

- BRUNS, R.; SONCIM, S. P.; DE SINAY-PHD, M. C. F. **Pesquisa operacional**: uma aplicação da teoria das filas a um sistema de atendimento. Instituto Militar de Engenharia. ENEGEP. 2001.
- CARRAVILLA, M. A.; OLIVEIRA, J. F. Operations research in agriculture: better decisions for a scarce and uncertain world. **Agris Online Papers in Economics and Informatics**. V. 5, Issue 2. p. 37-46. 2013.
- CAVALHEIRO, T. M. do N.; VAROLLO, M. G. F.; BERTOLUCI, E. A. Modelagem e Simulação computacional: um estudo envolvendo atendimento telefônico para cotações de frete de uma transportadora de JAHU. **Brazilian Journal of Technology**, v. 1, n. 2, p. 267-277, 2018.
- COOPER, J. CB. The poisson and exponential distributions. **Mathematical Spectrum**, v. 37, n. 3, p. 123-125, 2005.
- ENSSLIN, L.; VIANNA, W. B. O design na pesquisa quali-quantitativa em engenharia de produção—questões epistemológicas. **Revista Produção Online**, v. 8, n. 1, 2008.
- HEYER, R. **Understanding soft operations research**: the methods, their application and its future in the defence setting. 2004.
- HOU, C.; CAO, B.; FAN, J. A data-driven method to predict service level for call centers. **IET Communications**, v. 16, n. 10, p. 1241-1252. 2021.
- KABAK, Ö. et al. Efficient shift scheduling in the retail sector through two-stage optimization. **European Journal of Operational Research**, v. 184, n. 1, p. 76-90, 2008.
- KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. **The Annals of Mathematical Statistics**, p. 338-354, 1953.
- KIM, S. Conceptual design based on substance-field model in theory of inventive problem solving. **International Journal of Innovation, Management and Technology**, v. 3, n. 4, p. 306-309, 2012.
- LEE, A. M. Models of serial and cyclic processes. **Applied Queueing Theory**. Palgrave, London, p. 59-68. 1966.
- LEITE, L. de O.; REZENDE, D. A. **E-gov. estratégico: governo eletrônico para gestão do desempenho da administração pública**. Curitiba. Appris. 2017.
- LIMA, E. A.; MOLINARO, L. F. R. **Métodos soft da pesquisa operacional para estruturação de problemas em cenários complexos**. 2016.
- MARINS, C. S.; SOUZA, D. de O.; BARROS, M. da S. O uso do método de análise hierárquica (AHP) na tomada de decisões gerenciais—um estudo de caso. **XLI SBPO**, v. 1, p. 49, 2009.
- MORABITO, R.; LIMA, F. CR. Um modelo para analisar o problema de filas em caixas de supermercados: um estudo de caso. **Pesquisa Operacional**, v. 20, p. 59-71, 2000.

- PAMPLONA, L. de M. P.; LINS, M. P. E. A pesquisa operacional soft aplicada à governança para o desenvolvimento territorial. **Anais XIX Simpósio de Pesquisa Operacional & Logística da Marinha**. Publicação Online, V. 3, p. 81-91. 2020.
- RAUNER, M. S. et al. OR applied to health services: planning for the future with scarce resources. **European Journal of Operational Research**, v. 150, n. 1, p. 1-2, 2003.
- RUTKOWSKI, J. Qualidade no serviço público-um estudo de caso. **Gestão & produção**, v. 5, n. 3, p. 284-297, 1998.
- SANKARAN, S.; TAY, B. H.; ORR, M. Managing organizational change by using soft systems thinking in action research projects. **International Journal of Managing Projects in Business**, 2009.
- SANTOS, D. M. et al. Pesquisa operacional: um estudo de caso para minimização de custos em uma barbearia. **Gestão da Produção em Foco** V. 42, p. 64, 2021.
- SHORTLE, J. F. et al. **Fundamentals of queueing theory**. John Wiley & Sons, 2018.
- WANG, P. P.; GONG, M.; LEE, T. Waiting time in a combined first-come-first-served and shortest-time-first queue. **Computers & Operations Research**, v. 30, n. 11, p. 1727-1743, 2003.
- WILLIAMS, M. C.; GUNATUNGE, S. Composite case study demonstrating the use of soft systems methodology for analysing information management. In: **The Proceedings of the 17th International Conference of the Systems Dynamics Society and the Fifth Australian & New Zealand Systems Conference**, p. 20-23. 1999.
- YAN, X. Notice of retraction: the methodology of identifying and choosing customer values. **Anais International Conference on Management and Service Science**. IEEE, p. 1-5. 2009.