



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIENCIAS
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO

RODRIGO BARROS DE LIRA

**OTIMIZAÇÃO DA ALOCAÇÃO DE FUNCIONÁRIOS EM LOJA DE TELEFONIA
UTILIZANDO A TEORIA DAS FILAS E PROGRAMAÇÃO LINEAR**

RECIFE

2024

RODRIGO BARROS DE LIRA

**OTIMIZAÇÃO DA ALOCAÇÃO DE FUNCIONÁRIOS EM LOJA DE TELEFONIA
UTILIZANDO A TEORIA DAS FILAS E PROGRAMAÇÃO LINEAR**

Trabalho de Conclusão de Curso apresentado ao Curso de Engenharia de Produção da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Bacharel em Engenharia de Produção.

Área de concentração: Pesquisa Operacional.

Orientador: Alexandre Ramalho Alberti

RECIFE

2024

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Lira, Rodrigo Barros de.

Otimização da alocação de funcionários em loja de telefonia utilizando a teoria das filas e programação linear / Rodrigo Barros de Lira. - Recife, 2024. 54 p., tab.

Orientador(a): Alexandre Ramalho Alberti

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia de Produção - Bacharelado, 2024.

Inclui referências, apêndices.

1. Teoria das Filas. 2. Programação Linear. 3. Alocação de Funcionários. 4. Modelo M/M/k. 5. Atendimento ao Cliente. I. Alberti, Alexandre Ramalho . (Orientação). II. Título.

620 CDD (22.ed.)

RODRIGO BARROS DE LIRA

**OTIMIZAÇÃO DA ALOCAÇÃO DE FUNCIONÁRIOS EM LOJA DE TELEFONIA
UTILIZANDO A TEORIA DAS FILAS E PROGRAMAÇÃO LINEAR**

Trabalho de Conclusão de Curso
apresentado ao Curso de Engenharia de
Produção da Universidade Federal de
Pernambuco, como requisito parcial para a
obtenção do título de Bacharel em
Engenharia de Produção.

Aprovado em: 03/10/2024

BANCA EXAMINADORA

Prof. Dr. Alexandre Ramalho Alberti (Orientador)
Universidade Federal de Pernambuco

Prof^a. Dr^a. Isis Didier Lins (Examinador Interno)
Universidade Federal de Pernambuco

Prof. Dr. Cristiano Alexandre Virginio Cavalcante (Examinador Interno)
Universidade Federal de Pernambuco

RESUMO

O presente trabalho investiga a aplicação da Teoria das Filas e Programação Linear para otimizar a alocação de funcionários em uma loja de telefonia situada na região Nordeste do Brasil. A pesquisa visa determinar a quantidade ideal de colaboradores em função da demanda de clientes em diferentes horários do dia, utilizando o modelo M/M/k. A coleta de dados ocorreu ao longo de seis meses e incluiu informações sobre o fluxo de clientes, o tempo de atendimento e a variação de demanda. Com base nos resultados obtidos, foram desenvolvidas recomendações para o dimensionamento de funcionários e com a Programação Linear foi feita a elaboração dos turnos de trabalho, respeitando as regulamentações trabalhistas brasileiras e tendo impactos positivos sobre o atendimento ao cliente. A aplicação do modelo resultou em melhorias no tempo de espera dos clientes e eficiência operacional, além de propor uma redução do número de funcionários sem comprometer a qualidade do serviço. Por fim, a pesquisa oferece contribuições práticas, destacando a relevância da Teoria das Filas e da Programação Linear para a gestão de operações em ambientes de varejo.

Palavras-chave: Teoria das Filas, Programação Linear, Alocação de Funcionários, Modelo M/M/k, Atendimento ao Cliente.

ABSTRACT

This study investigates the application of Queueing Theory and Linear Programming to optimize employee allocation in a telecommunications store located in the Northeast region of Brazil. The research aims to determine the ideal number of employees based on customer demand at different times of the day, using the M/M/k model. Data collection took place over six months and included information on customer flow, service time, and demand variation. Based on the results obtained, recommendations for staff sizing were developed, and with Linear Programming, work shifts were created, adhering to Brazilian labor regulations and positively impacting customer service. The application of the model resulted in improvements in customer wait times and operational efficiency, as well as proposing a reduction in the number of employees without compromising service quality. Finally, the research offers practical contributions, highlighting the relevance of Queueing Theory and Linear Programming for operations management in retail environments.

Keywords: Queueing Theory, Linear Programming, Staff Allocation, M/M/k Model, Customer Service.

SUMÁRIO

1.	INTRODUÇÃO	7
1.1.	CONTEXTO	8
1.2.	PROBLEMA E JUSTIFICATIVA	9
1.3.	OBJETIVOS.....	10
1.4.	ESTRUTURA DO TRABALHO	11
2.	FUNDAMENTAÇÃO TEÓRICA E REVISÃO DA LITERATURA.....	13
2.1.	TEORIA DAS FILAS.....	13
2.1.1.	Notação.....	15
2.1.2.	Disciplina das filas.....	16
2.1.3.	Modelos de fila	17
2.1.3.1.	Filas de Espera com População Infinita e Canal de Atendimento Único (M/M/1)	18
2.1.3.2.	Filas de Espera com população infinita e canal de atendimento múltiplo (M/M/k).....	19
2.1.3.3.	Modelo M/Ek/s e M/Ek/1	21
2.2.	PROGRAMAÇÃO LINEAR.....	22
2.3.	REVISÃO DA LITERATURA	24
3.	ESTUDO DE CASO EM LOJA DE TELEFONIA.....	29
3.1.	ETAPAS DA PESQUISA	30
3.2.	RESULTADOS E DISCUSSÕES.....	31
3.2.1.	Escolha do Modelo de Fila M/M/k	31
3.2.2.	Validação da distribuição de taxa de chegada e resultados	32
3.2.2.1.	Análise do Fluxo de Clientes.....	33
3.2.2.2.	Performance do Modelo M/M/k	35
3.2.3.	Elaboração dos Turnos de Trabalho	40
3.2.3.1.	Considerações Preliminares	40
3.2.3.2.	Determinação do Número de Funcionários por Turno de Trabalho.....	41
4.	CONSIDERAÇÕES FINAIS	50
	REFERÊNCIAS	52
	APÊNDICE A – CÓDIGO TESTE DE KOLMOGOROV-SMIRNOV	54

1. INTRODUÇÃO

No dinâmico cenário do comércio varejista, onde a competição é acirrada e a satisfação do cliente é essencial, a eficiência operacional desempenha um papel fundamental. Segundo Bordoloi et al. (2023), a espera excessiva — ou até mesmo a mera expectativa de enfrentar longas filas — pode resultar em perda de vendas. Dessa forma, o gerenciamento eficaz das filas de clientes nas lojas, garantindo uma experiência positiva e maximizando a utilização dos recursos disponíveis, é imprescindível. Nesse contexto, a Teoria das Filas surge como uma ferramenta para analisar e otimizar esses processos, oferecendo insights que podem ser aplicados para melhorar tanto a qualidade do serviço quanto a eficiência.

No caso específico de uma loja de telefonia, o estudo da Teoria das Filas é importante devido a complexidade do atendimento e pela diversidade de serviços prestados, que incluem desde a compra de aparelhos até a resolução de questões técnicas e ajustes de planos de telefonia. O fluxo de clientes pode ser bastante irregular, com picos de demanda em horários e dias específicos, como durante lançamentos de novos produtos ou promoções. A capacidade de prever e gerenciar adequadamente essas variações de demanda é crucial para evitar longas esperas, que prejudicam a experiência do cliente e podem comprometer a imagem da empresa.

Além da Teoria das Filas, a Programação Linear desempenha um papel estratégico na otimização da alocação de recursos, como a definição do número ideal de funcionários em cada turno de atendimento. A Programação Linear permite modelar o problema de alocação de forma a atender as restrições operacionais, como a quantidade mínima de funcionários necessária em cada horário, a gestão de pausas para almoço e a carga horária máxima permitida por lei. A importância dessa ferramenta está na sua capacidade de encontrar soluções ótimas que minimizem custos e maximizem o uso eficiente dos recursos humanos, garantindo um nível adequado de serviço ao cliente em todos os momentos do dia.

O fenômeno central desta pesquisa é a gestão de filas em uma loja, uma realidade que muitos consumidores conhecem bem, especialmente em períodos de alta demanda, como feriados e promoções. Filas mal gerenciadas podem ser frustrantes para os clientes, resultando em insatisfação, perda de vendas e subutilização dos funcionários. Além dos

impactos psicológicos que a espera pode causar nos consumidores, Bordoloi et al. (2023) destacam a importância das duas "Leis do Serviço" para a gestão de filas. A primeira lei afirma que a satisfação do cliente depende da diferença entre suas expectativas e percepções: quando o serviço supera as expectativas, o cliente sai satisfeito, o que pode levar a recomendações positivas. A segunda lei ressalta que é difícil reverter uma primeira impressão negativa, já que essa experiência inicial molda a percepção subsequente. Portanto, reduzir o tempo de espera e tornar a experiência de fila mais agradável são fatores críticos para o sucesso.

A gestão de filas envolve diversos aspectos, como a alocação de caixas de atendimento, a criação de estratégias para redistribuir clientes entre diferentes filas e a previsão de tempos de espera. A Programação Linear auxilia na tomada de decisões operacionais, garantindo que os recursos sejam alocados de maneira eficiente, minimizando o tempo de espera e melhorando a satisfação do cliente.

O interesse pelo estudo da Teoria das Filas e da Programação Linear em uma loja de telefonia surgiu a partir das dificuldades operacionais observadas nesse ambiente. A necessidade de melhorar a qualidade do atendimento, otimizar o uso dos recursos e criar uma experiência de compra mais satisfatória impulsionou a escolha desse tema. A busca por soluções inovadoras e eficazes para a gestão de filas, que proporcionassem melhorias concretas tanto para a empresa quanto para os clientes, foi o que motivou esta pesquisa.

1.1. CONTEXTO

O problema a ser solucionado é definir a quantidade de funcionários necessários para uma das lojas de uma empresa de telefonia presente na região Nordeste de uma das principais empresas de telecomunicações do Brasil, utilizando a Teoria das Filas e Programação Linear.

A empresa em questão possui uma ampla gama de serviços que incluem telefonia fixa, telefonia móvel, banda larga e TV por assinatura. A empresa está presente em todos os estados do Brasil. Sua presença nacional permite que seus serviços sejam acessíveis para a maioria dos brasileiros.

Além disso, a empresa é reconhecida por seus investimentos em infraestrutura de rede e inovação tecnológica. Ela tem buscado expandir sua cobertura de banda larga, tanto

fixa quanto móvel, e tem se dedicado ao desenvolvimento de tecnologias avançadas para melhorar a conectividade em todo o país.

1.2. PROBLEMA E JUSTIFICATIVA

A empresa enfrenta dificuldades na definição do número ideal de funcionários em cada loja. É essencial que haja uma quantidade adequada de colaboradores para atender à demanda dos clientes, evitando longas filas e garantindo um serviço de qualidade. No entanto, também é necessário considerar fatores como o horário de almoço dos funcionários e a carga horária máxima permitida por lei.

Para solucionar esse problema, a empresa optou por aplicar tanto a Teoria das Filas quanto a Programação Linear, duas abordagens matemáticas complementares que visam otimizar a alocação de recursos em sistemas de atendimento. A Teoria das Filas se destaca por sua capacidade de modelar e prever o comportamento das filas, permitindo à empresa tomar decisões informadas sobre o dimensionamento da equipe, de acordo com as flutuações na demanda de clientes. Já a Programação Linear é crucial para garantir que os horários de trabalho e a alocação dos funcionários sejam otimizados com base em restrições como o horário de almoço e a carga horária máxima permitida por lei, respeitando as exigências operacionais e legais.

Com base nos dados sobre o fluxo de clientes em cada loja, a análise de padrões e a identificação dos horários de pico permitem determinar a quantidade mínima de funcionários necessária para atender à demanda nesses períodos críticos. A Programação Linear auxilia na criação de um cronograma que distribua eficientemente os colaboradores ao longo do dia, evitando sobrecargas ou ociosidade. Além disso, a redistribuição dos recursos durante o horário de almoço é fundamental para evitar uma queda na capacidade de atendimento nesse intervalo, o que é facilitado pelo uso da Programação Linear para ajustar turnos e maximizar a eficiência.

Outro fator crítico a ser considerado é o respeito à carga horária máxima de trabalho para cada funcionário, de acordo com a legislação trabalhista vigente. A Programação Linear torna possível criar horários que evitem o excesso de horas extras e a fadiga dos colaboradores, preservando sua saúde e garantindo conformidade com as leis trabalhistas, enquanto mantém o atendimento adequado ao cliente.

A pesquisa sobre a aplicação da Teoria das Filas e da Programação Linear nas operações de uma loja de telefonia apresenta uma relevância significativa no contexto prático. A implementação de estratégias baseadas nos resultados dessas análises pode gerar impactos positivos diretos na operação da loja. Uma alocação mais eficiente de recursos humanos, por exemplo, contribui para otimizar a distribuição dos funcionários em horários de maior demanda, o que reduz os tempos de espera dos clientes e melhora a qualidade do atendimento. Além disso, essa eficiência operacional se traduz em uma experiência mais satisfatória para o cliente, aumentando as chances de fidelização e elevando o desempenho geral da loja.

Os benefícios da pesquisa prometem desencadear uma série de melhorias operacionais. Ao otimizar a gestão de filas e os cronogramas de trabalho por meio da Programação Linear, a loja pode oferecer um atendimento mais ágil, aumentando a satisfação dos clientes e fortalecendo sua fidelidade. Além disso, a alocação estratégica dos recursos humanos pode reduzir custos operacionais e maximizar a eficiência financeira, resultando em um negócio mais competitivo. A aplicação dessas metodologias também contribui para o avanço das práticas empresariais e para o desenvolvimento acadêmico da Teoria das Filas e da Programação Linear, consolidando essas ferramentas como essenciais em diversos contextos.

A relevância da aplicação dessas abordagens matemáticas em cenários operacionais é demonstrada em estudos como o de Bahadori et al. (2014), que utilizaram a Teoria das Filas para otimizar a operação de uma farmácia hospitalar. No setor varejista, essas conclusões podem ser aplicadas para melhorar a eficiência operacional e a qualidade do atendimento ao cliente, influenciando positivamente toda a dinâmica da indústria. Isso reforça a importância da Teoria das Filas e da Programação Linear como ferramentas para a otimização e aprimoramento de processos operacionais em diversos setores, incluindo o varejo.

1.3. OBJETIVOS

O objetivo deste trabalho é desenvolver um modelo de alocação de funcionários baseado na Teoria das Filas para determinar a quantidade ideal de colaboradores em uma loja de telefonia da região Nordeste, garantindo um atendimento eficiente e de qualidade.

Os objetivos específicos são:

- Analisar o fluxo de clientes em diferentes horários na loja, identificando padrões e horários de pico.
- Utilizar os dados coletados para calcular a demanda média de atendimento em cada loja durante os períodos de maior movimento.
- Aplicar os princípios da Teoria das Filas para determinar a quantidade mínima de funcionários necessária para atender à demanda identificada.
- Determinar os turnos de trabalho e quantos funcionários são necessários em cada, utilizando a Programação Linear para respeitar as regulamentações trabalhistas quanto à carga horária máxima e evitando horas extras excessivas.
- Propor recomendações estratégicas à empresa com base nos resultados obtidos, visando a melhoria contínua do atendimento ao cliente e a otimização dos recursos humanos.

1.4. ESTRUTURA DO TRABALHO

O Capítulo 1 deste trabalho contém a introdução, onde são apresentados o contexto da loja no Nordeste brasileiro e os desafios relacionados à alocação eficiente de funcionários. A problemática deste estudo é expressa pela seguinte questão: Como otimizar a quantidade de colaboradores em função do fluxo de clientes, aplicando a Teoria das Filas? Este capítulo também apresenta a justificativa e a relevância do estudo, além de detalhar os objetivos gerais e específicos da pesquisa.

O Capítulo 2 aborda a fundamentação teórica e a revisão da literatura. Nele, são explorados os principais conceitos da Teoria das Filas e Programação Linear, juntamente com estudos e aplicações práticas em ambientes de serviços e outros setores que enfrentam desafios semelhantes na gestão de filas e atendimento ao cliente.

O Capítulo 3 descreve a modelagem e resultados deste trabalho. Ele começa detalhando os métodos de coleta de dados, como a observação dos fluxos de clientes e a cronometragem dos tempos de atendimento. A seguir, são discutidos os procedimentos de análise dos dados, utilizando modelos de Teoria das Filas para calcular a demanda e definir a quantidade mínima de funcionários necessária, sendo a quantidade mínima o input para o

problema de Programação Linear que define a quantidade de funcionários por turno. O capítulo também aborda as limitações e desafios enfrentados durante o processo de coleta e análise dos dados. Os resultados incluem a análise dos padrões de atendimento, a identificação de horários de pico e as recomendações para a alocação de funcionários. Este capítulo também discute as implicações desses resultados para a melhoria da eficiência operacional e da satisfação dos clientes.

O Capítulo 4 contém as considerações finais. Nele, são discutidas as principais conclusões do estudo, destacando como as estratégias sugeridas podem ser implementadas na loja de telefonia. Este capítulo também aborda as limitações encontradas durante a pesquisa e oferece sugestões para estudos futuros que possam expandir ou aprofundar a aplicação da Teoria das Filas em diferentes contextos.

2. FUNDAMENTAÇÃO TEÓRICA E REVISÃO DA LITERATURA

A primeira parte deste capítulo apresenta os fundamentos básicos sobre os quais a pesquisa apresentada neste trabalho foi realizada. A segunda parte, por sua vez, apresenta uma revisão de trabalhos acadêmicos que abordam temas relacionados ao assunto deste trabalho.

2.1. TEORIA DAS FILAS

Sistemas que possuem capacidade limitada de atendimento e lidam com demandas que chegam de forma aleatória estão sujeitos à formação de filas, dependendo das condições específicas observadas. A existência de filas pode causar atrasos no atendimento ou, mesmo quando os atrasos não ocorrem, o simples fato de esperar costuma ser percebido negativamente pelos clientes, seja em filas físicas ou em processos realizados remotamente por prestadores de serviço (WEISS; TUCKER, 2018). Nesse cenário, a Teoria das Filas surge como uma ferramenta fundamental para estudar e modelar matematicamente o comportamento desses sistemas. Sua aplicação abrange diversos contextos em que há espera e atendimento, sendo essencial para a análise e o planejamento de sistemas, com o objetivo de aprimorar tanto a qualidade do serviço quanto a eficiência operacional.

Belfiore e Favero (2013) afirmam que a Teoria das Filas se baseia em modelos que consideram a presença de um ou mais servidores responsáveis por prestar determinado tipo de serviço. Esses modelos são construídos com base em características específicas dos sistemas em estudo, tais como taxas de chegada dos clientes, tempos médios de serviço, capacidade dos servidores, entre outros.

A presença de filas é uma situação comum no cotidiano das pessoas, conforme destacado por Hillier e Lieberman (2006). Ao realizar compras ou obter serviços diversos, frequentemente nos deparamos com a necessidade de esperar em filas. Nesse sentido, a Teoria das Filas está diretamente relacionada à economia e desempenha um papel fundamental na qualidade de vida das pessoas.

Arenales et al (2007) ressaltam que a análise dos sistemas de filas busca oferecer tempos de espera e serviços satisfatórios aos usuários. Embora a Teoria das Filas não seja

exatamente uma técnica de otimização, ela permite a determinação do desempenho dos sistemas por meio de métricas como o tempo médio de espera e a produtividade do sistema. Essas informações são essenciais para a projeção da capacidade de atendimento do sistema e para a tomada de decisões relacionadas à sua operação.

A classificação de um sistema de filas depende do ponto de vista do usuário e do operador, como afirmado por Arenales et al (2007). Para os usuários, as medidas de interesse incluem o tempo médio de serviço e de espera, a probabilidade de o tempo de espera exceder o esperado, o número médio de pessoas na fila e a possibilidade de o tamanho da fila ser maior que o esperado. Já para o operador do sistema, as medidas envolvem a utilização média do sistema, a ocupação média dos servidores, o período médio de ocupação em atendimentos contínuos e a possibilidade de o período de atendimento ser maior que o esperado.

O processo de chegada dos usuários em um sistema de filas é um fator crucial a ser considerado, como mencionado por Taha (2008). Esse processo descreve o intervalo de tempo entre as chegadas sucessivas de usuários. Embora, em geral, o processo de chegada seja considerado constante ao longo do tempo e independente do número de usuários presentes, exceções podem ocorrer. Por exemplo, um usuário pode desistir de ingressar em uma fila devido ao tamanho excessivo ou quando o sistema atinge sua capacidade máxima.

A disciplina da fila, conforme destacado por Taha (2008), determina a ordem de seleção e atendimento dos clientes. Existem diferentes políticas de disciplina de filas, tais como o critério de "primeiro a chegar, primeiro a ser servido" ou "último a chegar, primeiro a ser servido". Além disso, pode haver seleção prioritária dos clientes com base em algum critério específico. A fonte geradora de clientes pode ser tanto finita quanto infinita, dependendo da limitação na chegada de clientes ou da existência de uma oferta abundante.

O número de servidores e a forma como o atendimento é realizado são aspectos cruciais na Teoria das Filas. Conforme destacado por Arenales et al (2007), os servidores podem ser representados por um único indivíduo, um grupo de pessoas ou máquinas, podendo realizar o serviço simultaneamente ou em sequência. Taha (2008) menciona que o serviço pode ser realizado por servidores em paralelo, em série ou em redes, sendo essencial descrever a quantidade de atendentes e sua disposição nas instalações.

A capacidade do sistema é outro fator-chave a ser considerado na Teoria das Filas. Conforme salientado por Hillier e Lieberman (2006), os modelos de filas permitem

estabelecer a forma de operação de um sistema de filas de maneira eficiente. Um sistema com capacidade operacional excessiva pode acarretar altos custos, enquanto um sistema com capacidade operacional insuficiente resulta em filas e tempos de espera excessivos, levando à insatisfação dos usuários. Portanto, a utilização de modelos de filas é essencial para equilibrar o tempo de espera e o custo do serviço.

2.1.1. Notação

Para compreender adequadamente os problemas relacionados às filas, é essencial definir algumas terminologias comumente utilizadas em todos os modelos abordados. Essas definições fornecem a base para o estudo e a análise dos sistemas de filas.

- **Clientes:** São os indivíduos ou entidades que chegam ao sistema buscando atendimento.
- **$N(t)$:** Representa o número de clientes presentes na fila em um instante específico t (com $t \geq 0$).
- **k :** Indica o número de atendentes ou canais de atendimento paralelos disponíveis no sistema.
- **λ :** Define a taxa média de chegada de novos clientes (isto é, o número esperado de chegadas por unidade de tempo) quando há n clientes no sistema.
- **μ :** Refere-se à taxa média de atendimento, isto é, o número de clientes que são atendidos em um determinado intervalo de tempo.
- **ρ :** representa o fator de utilização ou taxa de ocupação de um sistema de filas. Indica a proporção do tempo em que os servidores estão ocupados em relação ao tempo total disponível.
- **L_q :** Representa o número médio de clientes que aguardam na fila por atendimento.
- **L :** Refere-se ao número médio total de clientes no sistema, incluindo aqueles que estão aguardando e aqueles que estão sendo atendidos.
- **W_q :** Indica o tempo médio que um cliente espera na fila, excluindo o tempo de atendimento propriamente dito.
- **W :** Refere-se ao tempo total médio que um cliente passa no sistema, incluindo o

tempo de espera na fila e o tempo de atendimento.

- **$P_n(t)$** : Define a probabilidade de haver exatamente n clientes no sistema em um instante t , dado o número de clientes presentes no instante inicial 0.
- **P_n** : Refere-se à probabilidade de haver exatamente n clientes no sistema em um determinado momento.
- **P_0** : Indica a probabilidade de o sistema estar vazio, ou seja, sem clientes.

Conforme Hillier e Lieberman (2006), a validade dos modelos de filas depende da estabilidade do sistema. Um sistema é considerado estável quando, após um período suficiente, seu comportamento se torna independente tanto do estado inicial quanto do tempo decorrido. Nessa condição, a distribuição probabilística do estado do sistema permanece constante ao longo do tempo. Para que isso ocorra, é fundamental que a taxa de utilização do sistema não ultrapasse 1 ($\rho < 1$).

2.1.2. Disciplina das filas

Como mencionado anteriormente, a estrutura básica de uma fila merece destaque. Nesse contexto, a forma como os usuários são atendidos define a disciplina da fila. Segundo as classificações de Taha (2008), essa disciplina pode ser dividida em diferentes tipos. Um deles é o FIFO (first in, first out), onde o primeiro a chegar é o primeiro a ser atendido, sendo o ideal para situações como atendimento em bancos, supermercados e lojas. Essa disciplina é particularmente adequada em contextos onde o senso de justiça no atendimento é importante, garantindo que os clientes sejam atendidos na ordem em que chegam, sem preferência ou discriminação, o que faz com que seja amplamente utilizada em serviços que lidam diretamente com o público, como os casos citados. Outra categoria é o LIFO (last in, first out), em que o último a chegar é o primeiro a ser atendido, comumente aplicado em sistemas em que o item mais recente é mais fácil de ser recuperado, como em controle de estoques. Além disso, existe o atendimento aleatório, em que os clientes são atendidos conforme entram no sistema de filas, sem preocupação com a ordem de chegada. Outra possibilidade é o sistema com prioridades, que pode ser preemptivo ou não preemptivo, onde o atendimento ocorre conforme prioridades estabelecidas, como no caso de idosos,

grávidas e pessoas com deficiência, que são atendidos de acordo com a maior para a menor prioridade. Esse sistema é frequentemente usado para dar preferência a tarefas menores, que são concluídas rapidamente e retornam à fila de tarefas prontas para serem executadas.

2.1.3. Modelos de fila

De acordo com Marins (2011), antes de iniciar o estudo dos modelos de filas, é essencial apresentar a notação proposta por Kendall-Lee, amplamente adotada na descrição desses sistemas. A notação de Kendall descreve um modelo de filas através de uma sequência de símbolos separados por barras inclinadas, na forma $A/B/s/C/N/Z$. Cada um desses símbolos representa uma característica específica do sistema: o primeiro símbolo 'A' refere-se à distribuição de probabilidade dos intervalos entre chegadas; o segundo 'B' trata da distribuição de probabilidade dos tempos de serviço; o terceiro 's' indica o número de servidores (ou atendentes) disponíveis no sistema; o quarto 'C' refere-se à capacidade máxima do sistema (número máximo de clientes que podem estar no sistema, incluindo os em fila e os em atendimento); o quinto 'N' representa o tamanho da população de origem das demandas, que pode ser finita ou infinita; e o sexto 'Z' descreve a disciplina de atendimento, que define a ordem em que os clientes são atendidos (por exemplo, FIFO - First In, First Out).

Quando o sistema tem capacidade ilimitada, a população é considerada muito grande, e a disciplina de atendimento segue o modelo FIFO, os três últimos símbolos 'C', 'N' e 'Z' podem ser omitidos. Alguns dos símbolos mais comuns usados nesta notação incluem:

- **M:** Para processos de chegadas com distribuição de Poisson e tempos de serviço com distribuição exponencial.
- **GI:** Para processos de chegadas independentes com distribuição geral (não necessariamente Poisson).
- **Ek:** Para distribuição de Erlang de ordem k, aplicada aos tempos de serviço.
- **D:** Para tempos de serviço constantes, usados em modelos determinísticos.
- **G:** Para distribuições gerais dos tempos de serviço.

Essa notação permite a descrição de diversos modelos de filas de maneira padronizada e clara. Nos modelos a serem apresentados, predominam as simbologias "M"

para chegadas com distribuição Poisson e “Ek” para tempos de serviço com distribuição de Erlang. Sendo eles, M/M/1, M/M/k, M/Ek/s e M/Ek/1

2.1.3.1. Filas de Espera com População Infinita e Canal de Atendimento Único (M/M/1)

No modelo de filas M/M/1, temos um sistema fundamental que apresenta apenas um canal de atendimento, o que significa que existe apenas um servidor disponível para atender os clientes. A disciplina de atendimento adota a regra FIFO (First In, First Out), garantindo que o primeiro cliente a chegar seja o primeiro a ser atendido. Este modelo assume uma fila com capacidade infinita, o que permite que um número ilimitado de clientes aguarde pelo atendimento, e opera em um estado de equilíbrio, onde as características do sistema permanecem constantes ao longo do tempo. A análise desse modelo é essencial para entender o comportamento de sistemas de atendimento e suas dinâmicas de espera (HILLER; LIEBERMAN, 2006).

Além disso, presume-se que a população de clientes é infinita, o que significa que há sempre novos clientes chegando ao sistema, sem que o número total de clientes seja limitado. Nesse cenário, podemos usar sete equações fundamentais para analisar o desempenho do sistema e resolver problemas relacionados a esse tipo de fila.

A utilização do sistema, que é um dos principais parâmetros de interesse, é dada pela equação $\rho = \lambda/\mu$. Nesta fórmula, ρ representa a fração do tempo em que o sistema está ocupado, ou seja, a taxa de utilização; λ corresponde à taxa de chegada dos clientes, medida em unidades de tempo; e μ é a taxa de atendimento, ou seja, o número médio de clientes atendidos por unidade de tempo.

Como já dito anteriormente para que o sistema M/M/1 seja estável e funcione adequadamente, é necessário que a taxa de chegada seja menor do que a taxa de atendimento, ou seja, $\lambda/\mu < 1$. Isso garante que o sistema não fique sobrecarregado a longo prazo, pois a taxa de utilização ρ deve ser menor que 1. Se ρ for igual ou maior que 1, o sistema nunca atingirá um estado estável, e as filas crescerão indefinidamente.

As equações a seguir, que envolvem cálculos como o número médio de clientes na fila, o tempo médio de espera e outras métricas, são todas válidas dentro dessa condição de estabilidade. Esse modelo é amplamente utilizado em situações reais onde há apenas

um ponto de atendimento e os clientes chegam de forma aleatória, como em caixas de supermercado, guichês de atendimento, ou centrais de suporte.

Quadro 1 – Equações para modelo M/M/1

Indicadores	Equações
Probabilidade do sistema estar vazio	$P_0 = 1 - \frac{\lambda}{\mu}$
Número de clientes esperando na fila	$Lq = \frac{\lambda^2}{\mu(\mu - \lambda)}$
Número de clientes esperando no sistema	$L = \frac{\lambda}{(\mu - \lambda)}$
Tempo de espera na fila	$Wq = \frac{\lambda}{\mu(\mu - \lambda)}$
Tempo de espera no sistema	$W = \frac{\lambda}{(\mu - \lambda)}$

Fonte: O Autor (2024)

2.1.3.2. Filas de Espera com população infinita e canal de atendimento múltiplo (M/M/k)

O modelo M/M/k é utilizado para representar sistemas de filas que envolvem uma população infinita de clientes e múltiplos canais de atendimento, ou seja, mais de um atendente disponível. Nesse modelo, assim como no M/M/1, a disciplina de atendimento segue o princípio FIFO. No entanto, aqui o número de atendentes é k , o que significa que vários clientes podem ser atendidos simultaneamente. É importante ressaltar que os valores das taxas de chegada λ e de atendimento μ podem não ser constantes, dependendo da configuração do sistema.

Conforme Hillier e Lieberman (2006), o modelo M/M/k é fundamentado em algumas premissas importantes: assume-se que os tempos entre chegadas de clientes seguem uma distribuição exponencial, característica de um processo de Poisson, o que significa que as chegadas ocorrem de maneira aleatória e independente. Da mesma forma, os tempos de atendimento também seguem uma distribuição exponencial independente e idêntica entre

os servidores. O número de atendentes, representado por k , é um inteiro positivo que determina quantos canais de atendimento estão disponíveis no sistema.

Como no modelo M/M/1, para que o sistema permaneça estável, é necessário que a taxa de utilização, ρ , seja inferior a 1, para que o sistema seja estável, o que significa que a capacidade total de atendimento (considerando os k servidores) é suficiente para lidar com a taxa de chegada de clientes. A taxa de utilização é calculada pela fórmula $\rho = \lambda / (k\mu)$, onde:

- λ representa a taxa média de chegada de clientes ao sistema, medida em clientes por unidade de tempo;
- μ é a taxa de atendimento de um único canal, ou seja, o número médio de clientes que cada atendente consegue atender por unidade de tempo;
- k é o número de atendentes disponíveis no sistema;
- ρ é o fator de utilização, que indica a proporção de tempo em que os atendentes estão ocupados.

As equações que seguem para esse modelo incluem cálculos importantes, como o número médio de clientes na fila, o tempo médio de espera de um cliente antes de ser atendido, e outras métricas de desempenho. Esses cálculos são válidos sob a condição de estabilidade, garantindo que o sistema possa operar de maneira eficiente ao longo do tempo.

Esse modelo M/M/ k é bastante utilizado em cenários onde há múltiplos pontos de atendimento, como em hospitais, bancos e call centers, onde a chegada de clientes é aleatória e o tempo de atendimento varia, mas a distribuição de chegadas e atendimentos segue padrões probabilísticos conhecidos.

Quadro 2: Equações para modelo M/M/K

Indicadores	Equações
Probabilidade do sistema estar vazio	$P_0 = \frac{1}{\left[\sum_{n=0}^{k-1} \left(\frac{1}{n!} \right) \left(\frac{\lambda}{\mu} \right)^n \right] + \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k \frac{k\mu}{(k\mu - \lambda)}}$
Probabilidade que uma chegada tenha que esperar	$P_k = \left(\frac{1}{k!} \right) \left(\frac{\lambda}{\mu} \right)^k \frac{k\mu}{(k\mu - \lambda)} P_0$

Número de clientes esperando na fila	$Lq = \frac{\lambda \mu \left(\frac{\lambda}{\mu}\right)^k P_0}{(k-1)! (k\mu - \lambda)^2}$
Número de clientes esperando no sistema	$L = \frac{\lambda \mu \left(\frac{\lambda}{\mu}\right)^k}{(k-1)! (k\mu - \lambda)^2} P_0 + \frac{\lambda}{\mu}$
Tempo de espera na fila	$Wq = \frac{\mu \left(\frac{\lambda}{\mu}\right)^k P_0}{(k-1)! (k\mu - \lambda)^2}$
Tempo de espera no sistema	$W = \frac{\mu \left(\frac{\lambda}{\mu}\right)^k P_0}{(k-1)! (k\mu - \lambda)^2} + \frac{1}{\mu}$

Fonte: O Autor (2024)

2.1.3.3. Modelo M/Ek/s e M/Ek/1

Outro modelo de distribuição para os tempos de atendimento é a Distribuição de Erlang, assim denominada em homenagem a Agner Krarup Erlang, pioneiro da Teoria das Filas. A função de densidade probabilística desta distribuição é:

$$f(t) = \frac{(\mu k)^k}{(k-1)!} t^{k-1} e^{-k\mu t}, \quad \text{para } t \geq 0$$

Nesta distribuição, são utilizados dois parâmetros principais: μ , que corresponde à taxa média de atendimento, e k , que indica o grau de variabilidade nos tempos de atendimento em relação à média. Ambos devem ser valores estritamente positivos, com a particularidade de que k precisa ser um número inteiro.

De acordo com Hillier e Lieberman (2006), a Distribuição de Erlang é particularmente valiosa, pois se trata de uma ampla família de distribuições com dois parâmetros, que se restringe a valores não negativos. Assim, distribuições empíricas de tempos de atendimento podem frequentemente ser bem aproximadas por uma distribuição de Erlang.

Neste cenário, a distribuição exponencial surge como um caso particular da distribuição de Erlang, quando $k=1$. À medida que k aumenta até o infinito, a variabilidade diminui. Para qualquer valor de k , a média da distribuição é dada por $1/\mu$, a variância é $1/(k\mu^2)$ e a moda é $(k-1)/(k\mu)$. Com base nas estimativas empíricas de média e variância

dos tempos de atendimento, essas fórmulas permitem escolher um valor adequado de k que melhor se aproxime dos dados observados.

No caso específico do modelo M/Ek/1, em que o tempo de atendimento segue uma distribuição de Erlang com parâmetro k e existe um único atendente, são utilizadas determinadas equações para descrever o sistema:

Quadro 3: Equações para modelo M/M/K

Indicadores	Equações
Número de clientes esperando na fila	$Lq = \frac{1+k}{2k} \frac{\lambda^2}{\mu(\mu - \lambda)}$
Número de clientes esperando no sistema	$L = \lambda W$
Tempo de espera na fila	$Wq = \frac{1+k}{2k} \frac{\lambda}{\mu(\mu - \lambda)}$
Tempo de espera no sistema	$W = Wq + \frac{\lambda}{\mu}$

Fonte: O Autor (2024)

Segundo Hillier e Lieberman (2006) o modelo M/Ek/s, com múltiplos atendentes, é modelado como uma cadeia de Markov em tempo contínuo, em que cada cliente passa por k fases exponenciais individuais em vez de um único atendimento. No entanto, a solução genérica para o estado estável, considerando $\rho = \lambda/s\mu$, não é facilmente obtida, exigindo métodos numéricos avançados para a resolução de casos específicos.

Este modelo torna-se útil em sistemas que apresentam variação significativa nos tempos de atendimento, permitindo maior flexibilidade e precisão na modelagem de processos reais.

2.2. PROGRAMAÇÃO LINEAR

A programação linear é um dos modelos mais importantes dentro da modelagem matemática e da otimização combinatória, amplamente utilizada na pesquisa operacional e

em diversos setores empresariais e governamentais. Essa técnica matemática visa otimizar sistemas sujeitos a um conjunto de restrições, com o objetivo de maximizar ou minimizar uma função objetivo, como lucros, custos ou eficiência operacional (Goldbarg & Luna, 2005). A essência da programação linear reside na busca por soluções ótimas para problemas envolvendo recursos limitados, respeitando restrições de natureza linear, como capacidade de produção, demanda de mercado, ou outras limitações operacionais impostas pelo sistema.

De acordo com Hillier & Lieberman (2006), a programação linear é uma ferramenta fundamental na pesquisa operacional, desempenhando um papel crucial na resolução de problemas de alocação de recursos. Isso abrange desde a distribuição de produtos em diferentes mercados até a programação de produção em fábricas. O uso dessa técnica permite que organizações otimizem seus processos produtivos ou logísticos, tomando decisões estratégicas que visam o uso eficiente dos recursos disponíveis, sempre em conformidade com as restrições impostas pelo ambiente operacional.

Um aspecto chave da programação linear é sua aplicabilidade em uma variedade de campos, como finanças, logística, engenharia e, até mesmo, áreas como marketing e transporte (Arenales et al., 2007). Por exemplo, ela pode ser usada para otimizar a gestão de carteiras financeiras, o fluxo de mercadorias em uma cadeia de suprimentos, ou a alocação de equipamentos e pessoal em uma planta de produção. Dentro desse contexto, a programação inteira, uma variação da programação linear que trata de variáveis de decisão que assumem apenas valores inteiros, é amplamente utilizada em situações que requerem soluções discretas, como em problemas de alocação de recursos que envolvem números inteiros (Arenales et al., 2007).

Para que um problema possa ser formulado como um modelo de programação linear, ele precisa atender a certas características fundamentais. Conforme explicado por Goldbarg & Luna (2005), essas características incluem: proporcionalidade, que assegura que o impacto de cada variável na função objetivo ou nas restrições seja linear; não negatividade, onde as variáveis de decisão devem ser não negativas, garantindo que as soluções sejam viáveis e fisicamente possíveis; aditividade, que permite que a função objetivo e as restrições possam ser expressas como a soma dos componentes individuais de cada atividade; e separabilidade, que possibilita a análise separada dos custos ou benefícios associados a cada parte do problema.

Essas características são cruciais para garantir que os modelos de programação linear possam ser resolvidos por métodos computacionais específicos, como o método Simplex, que se baseia em técnicas algébricas para explorar o espaço de soluções viáveis até encontrar a solução ótima. A formulação geral de um problema de programação linear pode ser expressa de forma algébrica, onde se busca maximizar ou minimizar uma função linear sujeita a restrições lineares de desigualdade ou igualdade. Essa estrutura algébrica é amplamente utilizada em modelos práticos e é aplicável em diferentes contextos operacionais e estratégicos (Goldbarg & Luna, 2005).

A programação linear, portanto, não só é uma ferramenta teórica de grande importância, mas também uma técnica amplamente empregada em aplicações práticas, ajudando organizações a resolver problemas complexos de maneira eficiente e eficaz.

2.3. REVISÃO DA LITERATURA

A prestação de serviços, especialmente no setor de telecomunicações, segundo Farias (2021), apresenta desafios complexos de gerenciamento e organização, visto que fatores externos, como condições climáticas, infraestrutura externa, deslocamentos e localização, dificultam a estimativa precisa do tempo total de atendimento, demanda diária fixa e dimensionamento adequado das equipes de trabalho. Nesse contexto, Farias (2021) nota que a habilidade gerencial, métodos eficientes e processos bem definidos são essenciais para garantir recursos humanos motivados e focados na excelência do serviço, bem como a disponibilidade dos materiais necessários.

A Teoria das Filas se apresenta como uma ferramenta útil para estabelecer a operacionalidade do sistema de atendimento ao cliente. Farias (2021) percebe que, através do tratamento das informações e análise dos resultados, é possível estabelecer parâmetros para a gestão de servidores e clientes, garantindo a satisfação dos clientes em relação aos prazos de atendimento. Além disso, a Teoria das Filas permite compreender como fatores conhecidos e eventos não rotineiros, como a pandemia, podem afetar a demanda de serviços. Por exemplo, no setor de telecomunicações, durante os primeiros meses de 2020, houve um aumento significativo na demanda devido ao isolamento social, o que exigiu o acréscimo de equipes de ativação/reparo e reparo/qualidade.

Na área de Tecnologia da Informação (TI), a teoria de filas também desempenha um

papel importante no dimensionamento de servidores. O estudo de Zanardi (2020) propôs o dimensionamento de servidores na área de TI, visando avaliar as condições atuais do sistema e sugerir melhorias. Por meio de entrevistas e análise de dados, o estudo identificou as áreas que mais demandam serviços relacionados à TI são custos, recursos humanos e cobrança. A metodologia adotada permitiu ajustar os dados para um modelo M/M/s, utilizando softwares para facilitar a análise dos resultados. Os resultados obtidos por Zanardi (2020) indicaram que as condições atuais do sistema eram adequadas, mas sugestões de melhoria foram propostas, como treinamentos mais efetivos para a equipe do Help Desk e a implementação de um campo obrigatório no canal de atendimento para melhor descrição dos problemas pelos usuários.

Além disso, na área da saúde, estudos relacionados à Covid-19 também utilizaram a Teoria das Filas para avaliar a capacidade de atendimento do sistema de saúde. No estudo de Silva (2021) realizado na cidade de Uberaba, foi analisada a situação dos leitos disponíveis para o tratamento de pessoas infectadas pelo vírus. Os resultados demonstraram que o sistema hospitalar seria capaz de atender a demanda da cidade até certo ponto, mas entraria em colapso caso a taxa de chegada de pacientes excedesse um determinado limite. Essas informações tinham a finalidade de alertar sobre a necessidade de medidas de prevenção e isolamento, bem como a possibilidade de ampliação do número de leitos disponíveis para garantir o atendimento adequado à população.

O artigo escrito por Lantz (2016), propõe uma abordagem inovadora para medir a capacidade efetiva em um processo de triagem hospitalar usando a Teoria das Filas. O estudo analisa a relação entre taxas de chegada, tempos de espera e medidas de capacidade para derivar estimativas precisas. Os resultados indicam que a capacidade efetiva não é linearmente dependente do número de enfermeiros, e a gestão da capacidade deve ser ajustada à demanda flutuante. Isso tem implicações práticas para a alocação de recursos e tomada de decisão. O método pode ser aplicado em outros processos de atendimento, proporcionando uma alternativa confiável a métodos tradicionais.

Também em um contexto hospitalar o artigo de Wiler (2013), apresenta um novo modelo baseado na Teoria das Filas para prever o efeito de cenários de aglomeração de pacientes nas taxas de pacientes que deixam de ser atendidos em departamentos de emergência. O estudo valida o modelo usando dados reais e destaca a capacidade do modelo de considerar cenários diferentes e prever taxas de pacientes não atendidos. Isso

pode ajudar na gestão da capacidade e na alocação de recursos.

No estudo de Chen (2014) um modelo de filas foi apresentado para descrever interrupções de tráfego causadas por perturbações de veículos em estradas. O modelo considera a evolução do tamanho de filas de congestionamento e sua relação com a probabilidade de interrupção do tráfego. Os resultados mostram a utilidade do modelo para estimar probabilidades de interrupção e otimizar estratégias de controle de tráfego. O estudo destaca a capacidade do modelo de vincular comportamentos individuais de condução com fenômenos macroscópicos de tráfego.

O estudo de Lima (2016) analisou o uso da Teoria das Filas no setor bancário, mais especificamente em uma agência, utilizando dados coletados do sistema bancário para modelagem e análise dos resultados. Verificou-se que a distribuição exponencial, ou de Poisson, foi eficaz para representar as distribuições de chegadas dos clientes nas filas judicial, negocial e caixa. No entanto, apenas a prestação de serviço judicial pôde ser modelada pela distribuição de Poisson no tempo de atendimento bancário. Foram observadas diferenças significativas no tempo de espera na fila e no tempo médio de permanência no sistema, indicando sazonalidade nos atendimentos, com maior demanda nos últimos dois meses do ano. A fila negocial apresentou regularidade no tempo de espera, enquanto a fila caixa teve um tempo médio de espera razoável. A análise dos dados permitiu prever o comportamento da chegada dos clientes e proporcionou informações essenciais para o planejamento da distribuição de atendentes nas filas.

Nesse estudo de Morabito (2000), a Teoria das Filas foi aplicada para modelar o tempo médio de espera dos consumidores nos caixas de supermercados. Foram explorados três modelos: um modelo $M/M/m$ de fila única, onde m é o número de caixas em operação, outro modelo de m modelos $M/M/1$ paralelos e independentes, e um modelo Markoviano mais geral. Por meio de um estudo de caso em um supermercado da rede Jaú-Serve, localizada no interior de São Paulo, os resultados indicaram que o modelo Markoviano mais geral foi uma boa aproximação para o sistema, ao contrário dos outros dois modelos. No entanto, são necessários mais experimentos para avaliar a precisão desse modelo e sua validação em casos práticos.

O estudo conduzido por Bahadori et al. (2014) abordou a otimização do desempenho de uma farmácia hospitalar, com foco na gestão eficiente da sua operação. Em um contexto no qual a farmácia hospitalar desempenha um papel crucial na administração e controle dos

processos de uso de medicamentos, os autores buscaram desenvolver uma abordagem inovadora. Através da aplicação da Teoria das Filas e da técnica de simulação, o estudo investigou a melhoria da eficiência e qualidade dos serviços prestados pela farmácia. Para alcançar esse objetivo, os pesquisadores realizaram uma análise detalhada dos fluxos de pacientes em dois turnos - manhã e tarde - em um hospital militar no Irã. A partir da coleta de dados, incluindo taxa de chegada e taxa de serviço, foram calculados indicadores-chave de desempenho da rede de filas. A simulação revelou informações importantes, como a possibilidade de redução dos tempos médios de espera e do número de pacientes na fila, por meio de ajustes na alocação de recursos e na organização dos processos. As conclusões do estudo ressaltaram a importância de estratégias como a multitarefa do pessoal e a realocação eficiente de recursos para etapas críticas, apontando para uma gestão mais eficaz da farmácia hospitalar, o que pode resultar em benefícios significativos para pacientes e funcionários.

Com relação a programação linear, Yi e Wang (2017) trata da criação de um modelo de programação linear inteira mista para otimizar cronogramas de trabalho-descanso em canteiros de obras sob altas temperaturas. Os autores apontam que trabalhadores da construção civil enfrentam desafios ergonômicos significativos, exacerbados pelo calor, o que impacta negativamente a produtividade e a segurança. O modelo desenvolvido visa maximizar o tempo produtivo total dos trabalhadores, levando em consideração fatores fisiológicos, o ambiente de trabalho e as regulamentações governamentais sobre períodos mínimos de descanso. Em experimentos numéricos, o modelo proposto apresentou um desempenho superior ao cronograma padrão utilizado em até 10%, o que sugere uma economia de custos significativa para o setor da construção.

No contexto da otimização, Yi e Wang destacam que a programação matemática oferece vantagens sobre heurísticas baseadas em simulação, como a obtenção de soluções ótimas que maximizam a produtividade. A modelagem considera a recuperação física dos trabalhadores durante os períodos de descanso e sugere cronogramas que atendem às exigências regulatórias sem comprometer a segurança. O estudo também propõe a utilização de técnicas de linearização para resolver o modelo de forma eficiente, utilizando solucionadores como o CPLEX.

Em termos de Programação Linear, Yi e Wang (2017) propõem um modelo de programação linear inteira mista para otimizar cronogramas de trabalho em canteiros de

obras sob altas temperaturas. O estudo demonstra que a aplicação da programação matemática pode trazer soluções ótimas e maximizar a produtividade, ao mesmo tempo em que garante a segurança dos trabalhadores e o cumprimento de regulamentações governamentais.

A presente pesquisa contribui para o campo ao aplicar a Teoria das Filas e a Programação Linear no contexto de uma loja de telefonia, focando na otimização da alocação de funcionários para maximizar a eficiência do atendimento ao cliente. De forma semelhante aos estudos citados, o trabalho utiliza modelos M/M/k para definir a quantidade ideal de servidores em cada turno, levando em consideração a demanda e o tempo médio de espera. No entanto, a principal diferença está na aplicação integrada de ambas as abordagens – filas e programação linear – para criar um cronograma de trabalho eficiente, considerando restrições trabalhistas e a necessidade de garantir folgas regulares para os funcionários, o que é uma inovação no contexto de lojas de telefonia. Assim, este trabalho oferece uma solução prática e adaptada às necessidades do setor, destacando-se pelo uso combinado de duas ferramentas poderosas de otimização.

3. ESTUDO DE CASO EM LOJA DE TELEFONIA

Este capítulo descreve o estudo de caso feito para otimizar a alocação de funcionários em uma loja de telefonia, assim como os resultados obtidos da otimização.

Com relação a metodologia, a presente pesquisa tem como objetivo investigar a aplicação da Teoria das Filas em uma loja de shopping específica, utilizando uma abordagem dedutiva. A abordagem dedutiva é escolhida porque busca estabelecer conclusões a partir de premissas verdadeiras (Lakatos & Marconi, 1991). Nesse sentido, a análise será focada em uma única loja, permitindo uma compreensão detalhada dos processos de atendimento ao cliente e da utilização dos recursos humanos.

O procedimento adotado será um estudo de caso, pois essa metodologia permite um detalhamento minucioso da situação investigada (Gil, 2008). Ao utilizar dados e informações específicas da loja em análise, o estudo de caso oferecerá uma perspectiva aprofundada da dinâmica das filas na área de atendimento ao cliente. Essa abordagem fornece insights valiosos sobre a eficiência operacional da loja e facilitará a identificação de possíveis melhorias nos serviços prestados.

Quanto à natureza, esta pesquisa é classificada como aplicada, uma vez que se baseia em conhecimentos práticos específicos para solucionar problemas reais (Gil, 2008). A abordagem aplicada é fundamental para extrair informações valiosas da realidade da loja de shopping em questão, permitindo a aplicação dos resultados obtidos para melhorar a eficiência e o desempenho dos serviços oferecidos.

Com relação aos objetivos, este estudo é caracterizado como descritivo (Gil, 1988). O foco principal será avaliar o desempenho e a demanda da loja por meio da medição de variáveis relacionadas ao atendimento ao cliente. A pesquisa descritiva é adequada para esse estudo, pois busca descrever os fenômenos observados e suas relações por meio de variáveis, fornecendo uma visão detalhada do contexto em que a Teoria das Filas é aplicada.

O estudo foi conduzido com uma abordagem quantitativa, utilizando a teoria de filas para determinar os parâmetros do modelo probabilístico que regem o atendimento de demandas na loja (Silva; Menezes, 2005). A abordagem quantitativa é escolhida devido à sua capacidade de quantificar dados e informações, possibilitando análises estatísticas que permitem uma compreensão mais precisa e objetiva dos fenômenos em estudo.

No que diz respeito à modelagem do problema de pesquisa, esta investigação adota

uma abordagem de modelagem de sistemas de fila, com dados de 6 meses do fluxo da loja em questão. A modelagem de sistemas de fila é uma técnica amplamente reconhecida para analisar o fluxo de chegadas e partidas em processos de atendimento, permitindo a representação matemática da interação entre clientes e recursos (Gross; Harris, 1985). Ao aplicar essa abordagem, pretende-se construir um modelo que simule o comportamento das filas na loja de shopping, considerando fatores como taxa de chegada de clientes, tempo médio de atendimento e número de servidores disponíveis.

3.1. ETAPAS DA PESQUISA

As etapas da pesquisa seguiram um fluxo lógico e sistemático, visando garantir a adequação dos métodos à problemática abordada. Na Etapa 1, foi realizada uma revisão bibliográfica sobre a Teoria das Filas e a Programação Linear. Essa etapa foi fundamental para embasar o trabalho teoricamente, fornecendo subsídios para a aplicação dos modelos matemáticos de filas no contexto da alocação de funcionários. Em seguida, na Etapa 2, foi realizada a coleta de dados preliminares relativos ao fluxo de clientes e horários de pico na loja de telefonia em análise. Esses dados serviram como base para a modelagem subsequente.

Na Etapa 3, os dados coletados foram analisados detalhadamente, permitindo a identificação de padrões no comportamento dos clientes ao longo do dia e da semana. Essa análise forneceu insights importantes para a adequação do modelo de filas à realidade observada na loja. A Etapa 4 envolveu o cálculo da demanda média de atendimento, seguido da aplicação dos princípios da Teoria das Filas, com o objetivo de determinar a quantidade mínima de funcionários necessária para garantir um atendimento eficiente. Nesta etapa, modelos como $M/M/k$ foram empregados para lidar com a variabilidade do fluxo de clientes e do tempo de serviço, na seção 3.2 será mais detalhado o motivo da escolha do modelo $M/M/k$.

Após a determinação das necessidades de pessoal, na Etapa 5, foram elaborados os turnos de trabalho, com a Programação Linear, considerando os resultados obtidos a partir da análise das filas e respeitando as regulamentações trabalhistas vigentes. Finalmente, na Etapa 6, foram realizadas simulações baseadas no modelo desenvolvido, e os resultados foram analisados. Ajustes finais foram feitos no modelo, visando garantir sua aplicabilidade prática e a melhoria contínua da alocação de funcionários na loja.

3.2. RESULTADOS E DISCUSSÕES

Este tópico tem como objetivo apresentar e discutir os resultados obtidos a partir da aplicação da Teoria das Filas no contexto de otimização da alocação de funcionários em uma loja de telefonia. Utilizando dados coletados ao longo de seis meses, os resultados fornecem uma visão detalhada sobre o fluxo de clientes, os tempos de espera, a utilização dos funcionários e a eficiência do atendimento nos diferentes horários de operação da loja.

Inicialmente, os dados brutos sobre a taxa de chegada de clientes e o tempo médio de atendimento foram analisados para identificar padrões de demanda e definir os horários de pico. Com base nessas informações, foram feitas inferências para identificar o modelo ideal de fila a ser aplicado para determinar a quantidade ideal de funcionários em cada período do dia, considerando os diferentes níveis de fluxo de clientes.

3.2.1. Escolha do Modelo de Fila M/M/k

Para modelar o sistema de atendimento da loja, foi escolhido o modelo de fila única com múltiplos atendentes (M/M/k). Este modelo foi selecionado por representar adequadamente o contexto da loja, onde vários atendentes (ou servidores) estão disponíveis para atender os clientes que chegam de maneira aleatória. As características da loja são compatíveis com as hipóteses do modelo M/M/k, que requerem:

- Uma população de clientes teoricamente infinita (não há limite máximo de clientes no shopping);
- Vários canais de atendimento operando de forma independente e simultânea (cada funcionário atende um cliente por vez);
- Fila única onde os clientes são atendidos pela ordem de chegada (disciplina de fila FIFO);
- As chegadas de clientes seguem uma distribuição de Poisson;
- Os tempos de atendimento seguem uma distribuição Exponencial.

No contexto da loja, a taxa de chegada de clientes foi estimada com base nos dados de fluxo coletados durante seis meses. O tempo médio de atendimento foi fornecido pela gerência, que verificou que esse tempo é de em média 20 minutos. Para fins de cálculo,

considerou-se que o tempo médio de atendimento segue uma distribuição exponencial. A gerência também forneceu o dado de que a loja deve manter um tempo total médio de permanência do cliente na loja (incluindo o atendimento e o tempo na fila) de 30 minutos. Dessa forma, o tempo médio de espera na fila deve ser de 10 minutos. Com os dados coletados, foi possível estimar tanto a taxa média de chegada de clientes (λ) e a capacidade de atendimento de cada funcionário (μ) que de acordo com o tempo médio fornecido, é igual a 3, sendo ambos os parâmetros fundamentais para a aplicação do modelo M/M/k.

Com base nos resultados obtidos a partir deste modelo, foi possível identificar o número ideal de funcionários que a loja precisa em horários de pico, otimizando a eficiência do atendimento. As análises realizadas fornecerão uma base para ajustes operacionais e melhorias contínuas no serviço oferecido pela loja.

3.2.2. Validação da distribuição de taxa de chegada e resultados

O processo de validação da taxa de chegada foi feito com base nos conceitos apresentados no referencial teórico, com o objetivo de identificar possíveis diferenças e propor melhorias que estejam alinhadas aos objetivos da pesquisa. Para isso, realizamos testes para verificar qual distribuição de probabilidade melhor se ajusta aos dados observados. O teste de Kolmogorov-Smirnov foi utilizado, com um nível de significância de 5% ($\alpha = 0,05$).

Os resultados mostraram que a hipótese nula, que supõe que os dados seguem uma distribuição de Poisson, não foi rejeitada. Isso significa que, de acordo com os resultados obtidos de valor-p igual a 0,2883 e estatística K-S igual a 0,2513, não há evidências suficientes para afirmar que os dados se desviam dessa distribuição. O valor-p maior que 0,05 indica que a diferença entre os dados observados e a distribuição de Poisson não é significativa. Em outras palavras, os dados se ajustam bem à distribuição de Poisson, sugerindo que o processo de chegada das demandas pode ser considerado homogêneo.

As análises foram feitas utilizando a linguagem Python, e os códigos desenvolvidos serão apresentados no apêndice deste trabalho.

3.2.2.1 Análise do Fluxo de Clientes

Os dados coletados foram organizados de acordo com a taxa média de chegada de clientes por hora (λ), conforme apresentado na tabela abaixo. Estes valores representam o número médio de clientes que chegam à loja por hora ao longo dos seis meses de observação.

Tabela 1: Taxa Média de Chegada

Período	Taxa Média de Chegada (λ)
09h – 10h	10,237
10h – 11h	14,593
11h – 12h	15,093
12h – 13h	17,381
13h – 14h	17,229
14h – 15h	18,280
15h – 16h	19,907
16h – 17h	19,712
17h – 18h	18,305
18h – 19h	16,559
19h – 20h	15,525
20h – 21h	11,627
21h – 22h	5,102

Fonte: O Autor (2024)

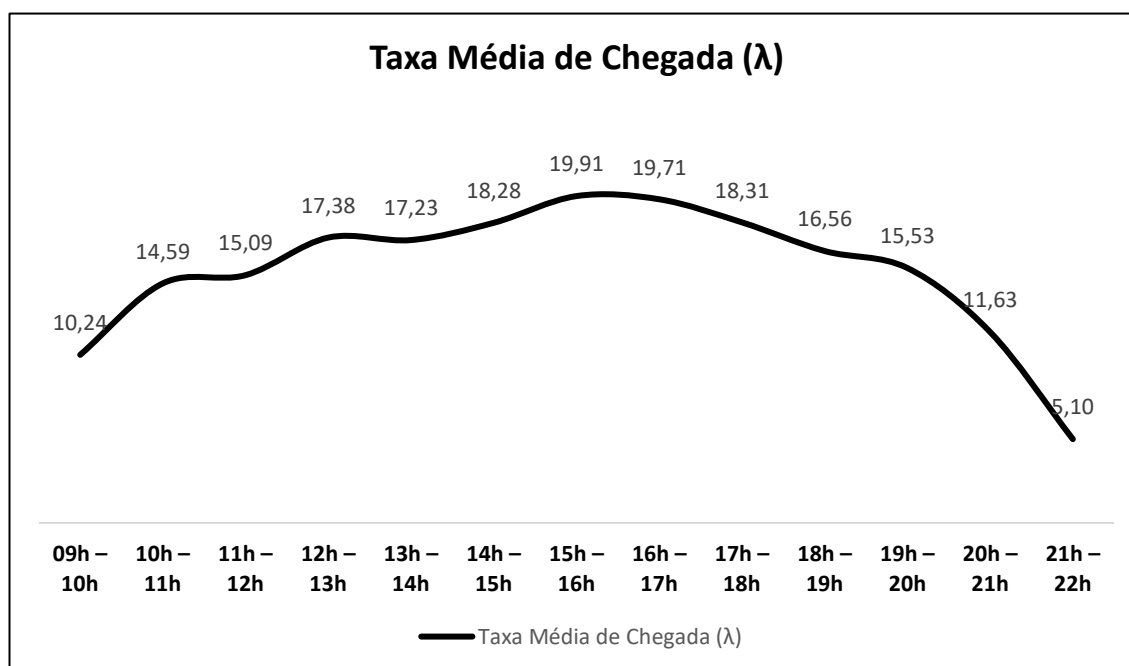
Com base nos dados acima, podemos observar que a demanda atinge o seu pico entre as 14h e as 18h, com uma taxa de chegada de clientes superior a 18 clientes por hora, atingindo o valor máximo de 19,91 clientes/hora às 15h. Esses horários representam os momentos de maior movimento, exigindo um número adequado de funcionários para evitar longas filas e garantir a qualidade do atendimento.

Os horários com menor fluxo de clientes foram observados durante o fim do dia, especialmente às 21h, com valores abaixo de 10 cliente por hora. Nesse horário, uma quantidade reduzida de funcionários é suficiente para manter o atendimento adequado.

Os padrões de demanda mostram um aumento gradual no número de clientes a partir das 10h, com um crescimento mais acentuado entre 12h e 15h. Após esse período de alta demanda, observa-se uma diminuição no fluxo de clientes a partir das 18h, com uma queda significativa após as 20h.

No gráfico abaixo, a curva representa a taxa média de chegada de clientes ao longo do dia, evidenciando os horários de pico e as quedas de fluxo em momentos de menor movimento.

Gráfico 1: Taxa Média de Chegada



Fonte: O Autor (2024)

Este gráfico permite identificar claramente os momentos de maior demanda, facilitando a tomada de decisões estratégicas sobre a alocação de funcionários em cada período.

Com base nos resultados desta análise, é possível recomendar um aumento no número de funcionários disponíveis durante os horários de pico, especialmente entre 14h e 18h, quando a demanda é mais intensa. Nos horários de menor fluxo, como pela manhã e após as 20h, a quantidade de colaboradores pode ser reduzida, otimizando os recursos humanos.

e evitando ociosidade.

Para implementar essas recomendações de forma eficaz, serão necessários turnos de trabalho que levem em consideração tanto a variabilidade na demanda quanto as exigências legais sobre carga horária e intervalos de descanso. O próximo passo consiste em aplicar o modelo M/M/k para determinar a quantidade exata de funcionários necessária em cada horário, buscando minimizar o tempo de espera dos clientes e melhorar a eficiência do atendimento.

3.2.2.2 Performance do Modelo M/M/k

Aqui serão apresentados os resultados da aplicação do modelo de filas M/M/k em diferentes períodos operacionais da loja, abrangendo desde horários de pico até momentos de baixo fluxo de clientes. A análise busca a maximização da eficiência do atendimento. A performance do modelo será avaliada em termos de tempo médio de espera, probabilidade de clientes na fila e utilização dos servidores.

Com base na coleta de dados sobre o fluxo de clientes ao longo de seis meses, foi possível determinar as taxas médias de chegada (λ) e atendimento (μ) da loja para cada hora do dia. O modelo M/M/k foi então aplicado, utilizando essas taxas como base para definir o número ideal de servidores (funcionários) necessários em cada período. Esse modelo foi escolhido porque considera múltiplos servidores atendendo a uma fila única, o que reflete a realidade operacional da loja.

A seguir, os resultados da aplicação do modelo serão divididos em três categorias de análise: horários de pico, horários normais e horários de baixo fluxo. Para cada período, serão apresentados os principais indicadores de desempenho, incluindo o número ideal de servidores, o tempo médio de espera na fila (W_q), a probabilidade do sistema estar vazio (P_0), e a utilização dos servidores (ρ). O número ideal de servidores é definido quando o tempo médio de espera na fila é menor que 10 minutos.

Nos horários de pico, o fluxo de clientes atinge seus valores máximos, exigindo um maior número de funcionários para manter um nível satisfatório de atendimento. No estudo, os horários de pico foram identificados entre 14h e 18h, quando a taxa de chegada de clientes (λ) chega a mais de 18 chegadas por hora.

Tabela 2: Resultados Modelo M/M/K em horários de pico

Período	Taxa de Chegada (λ)	Taxa de Atendimento (μ)	Número Ideal de Servidores (k)	Tempo Médio de Espera (W_q)	Probabilidade de o Sistema estar vazio (P_0)	Utilização dos Servidores (ρ)
14h – 15h	18,28	3	8	4,0 min	0,19%	76,2%
15h – 16h	19,91	3	8	7,7 min	0,10%	82,9%
16h – 17h	19,71	3	8	7,1 min	0,10%	82,1%
17h – 18h	18,3	3	8	4,0 min	0,19%	76,3%

Fonte: O Autor (2024)

Durante os horários de pico, o modelo M/M/k sugere a necessidade de 8 servidores, nos horários de pico. Como pode ser observado, o tempo médio de espera dos clientes (W_q) permanece abaixo de 10 minutos, mesmo nos períodos de maior fluxo. Isso demonstra que o modelo é eficaz em garantir uma boa qualidade de atendimento, mesmo durante os períodos mais críticos.

A probabilidade de o sistema estar vazio (P_0) durante esses horários varia entre 0,10% e 0,19%, o que é relativamente baixo devido à alta demanda. Entretanto, a utilização dos servidores (ρ) se mantém em níveis satisfatórios, entre 76% e 83%, garantindo que os funcionários estejam ocupados sem estarem sobrecarregados, evitando a exaustão e mantendo um equilíbrio no serviço prestado.

Os horários normais são caracterizados por uma demanda moderada, ocorrendo, por exemplo, entre 10h e 13h, e novamente entre 18h e 21h. Nessas faixas, a taxa de chegada de clientes é significativamente menor que nos horários de pico, mas ainda exige uma quantidade adequada de funcionários para manter o fluxo de atendimento eficiente.

Tabela 3: Resultados Modelo M/M/K em horários normais

Período	Taxa de Chegada (λ)	Taxa de Atendimento (μ)	Número Ideal de Servidores (k)	Tempo Médio de Espera (Wq)	Probabilidade do Sistema estar vazio (P0)	Utilização dos Servidores (ρ)
10h – 11h	14,59	3	6	9,5 min	0,56%	81,07%
11h – 12h	15,09	3	7	3,4 min	0,58%	71,87%
12h – 13h	17,38	3	7	9,0 min	0,22%	82,77%
13h – 14h	17,23	3	7	8,2 min	0,23%	82,04%
18h – 19h	16,56	3	7	6,2 min	0,32%	78,85%
19h – 20h	15,53	3	7	4,0 min	0,49%	73,93%

Fonte: O Autor (2024)

Nos horários normais, o modelo M/M/k sugere de 6 a 7 servidores como ideal para manter o tempo médio de espera abaixo de 10 minutos. Isso garante que os clientes sejam atendidos de maneira rápida e eficiente, mantendo a qualidade do serviço. A probabilidade de o sistema estar vazio (P0) é maior em relação aos horários de pico, variando entre 0,22% e 0,58%. A utilização dos servidores (ρ) também é mais equilibrada, permanecendo entre 71% e 83%.

Os horários de baixo fluxo ocorrem, por exemplo, entre 9h e 10h e a partir das 21h, quando a demanda é consideravelmente menor. Durante esses períodos, é possível operar com um número mínimo de servidores, mantendo a eficiência e o controle sobre os recursos.

Tabela 4: Resultados Modelo M/M/K em horários de baixo fluxo

Hora	Taxa de Chegada (λ)	Taxa de Atendimento (μ)	Número Ideal de Servidores (k)	Tempo Médio de Espera (Wq)	Probabilidade do Sistema estar vazio (P0)	Utilização dos Servidores (ρ)
09h – 10h	10,24	3	5	4,4 min	3,89%	68,25%
20h – 21h	11,63	3	5	9,0 min	1,56%	77,51%
21h – 22h	4,92	3	3	4,8 min	16,56%	56,69%

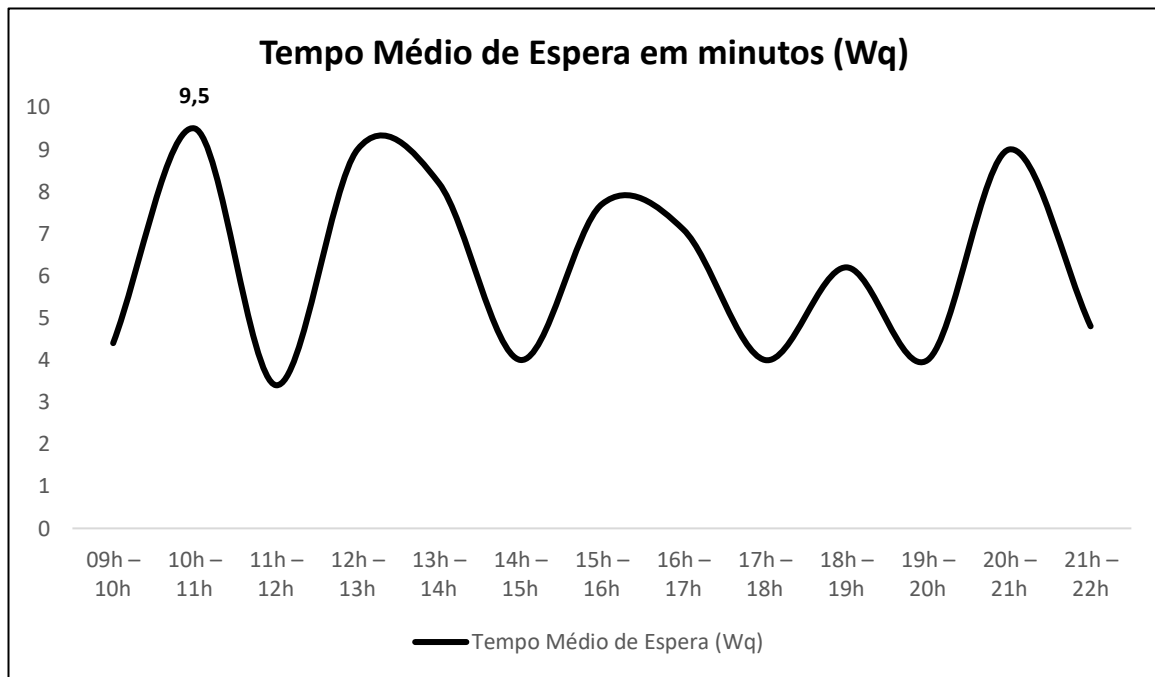
Fonte: O Autor (2024)

Durante os horários de baixo fluxo, o modelo sugere a utilização de 3 a 5 servidores. A probabilidade de o sistema estar vazio (P0) é mais alta que nos demais horários, variando entre 1,5% e 17%, o que reflete a baixa demanda nesses horários. A utilização dos servidores (ρ) é também é mais baixa, indicando que os funcionários têm menos pressão durante esses períodos.

A aplicação do modelo M/M/k mostrou-se eficaz em equilibrar a demanda e a oferta de atendimento em todos os períodos do dia. Durante os horários de pico, a utilização dos servidores é otimizada, com tempo de espera controlado e atendimento eficiente. Nos horários normais, o modelo garante a presença de uma quantidade adequada de funcionários, mantendo o tempo de espera dentro de limites aceitáveis. Nos horários de baixo fluxo, a redução do número de servidores evita ociosidade e garante que os recursos sejam utilizados de maneira mais eficiente.

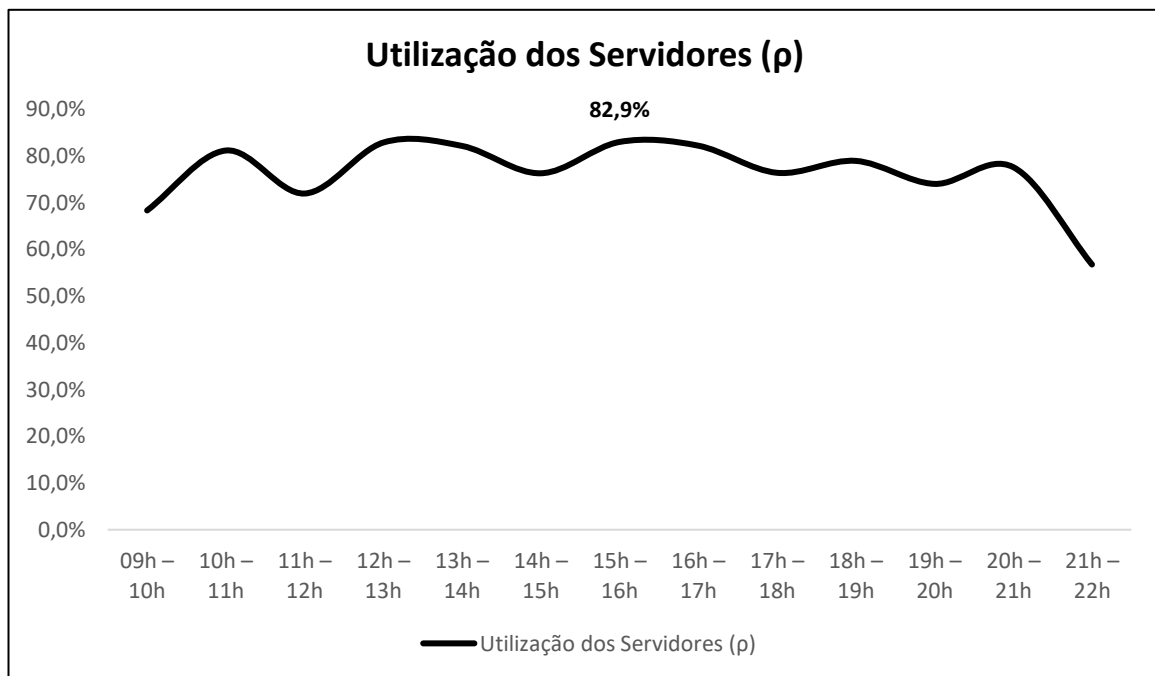
Os gráficos a seguir ilustram o desempenho do modelo M/M/k em termos de tempo médio de espera (Wq) e utilização dos servidores (ρ) em diferentes horários do dia, proporcionando uma visualização clara dos resultados obtidos.

Gráfico 1: Tempo Médio de Espera



Fonte: O Autor (2024)

Gráfico 2: Utilização dos Servidores

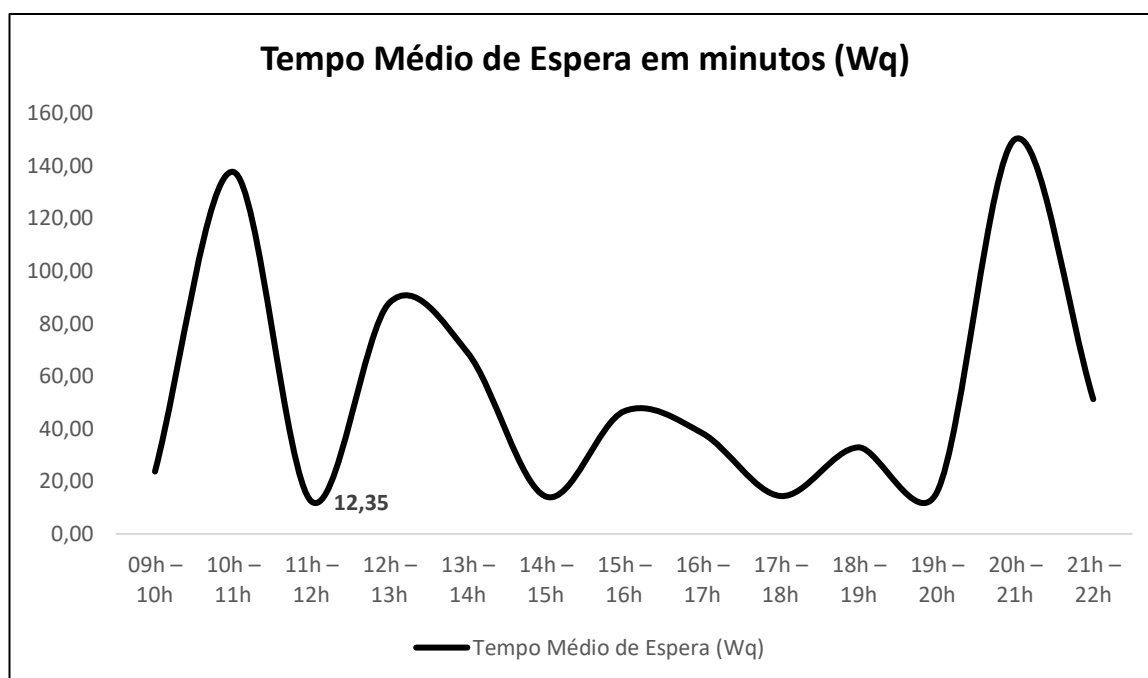


Fonte: O Autor (2024)

Para fins de comparação, o gráfico abaixo mostra como a redução de um funcionário em cada intervalo de tempo resultaria em um aumento significativo no tempo de espera na

fila. Em todos os casos, o tempo de espera ultrapassaria os 10 minutos estabelecidos como limite.

Gráfico 3: Tempo Médio de Espera com um funcionário a menos



Fonte: O Autor (2024)

Esses resultados fornecem uma base sólida para futuras decisões estratégicas sobre a alocação de recursos humanos na loja, permitindo que a empresa otimize sua operação e melhore a experiência do cliente.

3.2.3. Elaboração dos Turnos de Trabalho

A partir dos cálculos realizados com o modelo M/M/k, foi possível determinar o número ideal de funcionários para atender adequadamente à demanda em diferentes horários. Este tópico apresenta a elaboração dos turnos de trabalho que otimizam os recursos humanos, considerando as regulamentações trabalhistas brasileiras, como carga horária, intervalos para descanso e refeições, além das horas extras.

3.2.3.1 Considerações Preliminares

De acordo com a Consolidação das Leis do Trabalho (CLT), a jornada máxima

permitida é de 44 horas semanais, com limite de 8 horas diárias. Cada jornada que ultrapasse 6 horas deve incluir um intervalo de pelo menos 1 hora para almoço ou descanso.

3.2.3.2. Determinação do Número de Funcionários por Turno de Trabalho

A análise do modelo M/M/k mostrou que a demanda de funcionários varia ao longo do dia, com picos entre 14h e 18h. Para garantir atendimento eficiente sem sobrecarga, foi construído um modelo que, posteriormente, foi analisado no Excel utilizando o software Solver e a metodologia LP Simplex. Além do número ideal de funcionários, foi considerada a necessidade de três colaboradores no período das 8h às 9h para organizar e realizar a abertura da loja.

Tabela 5: Número ideal de Servidores por Hora

Período	Número Ideal de Servidores (k)
08h – 09h (n8)	3
09h – 10h (n9)	5
10h – 11h (n10)	6
11h – 12h (n11)	7
12h – 13h (n12)	7
13h – 14h (n13)	7
14h – 15h (n14)	8
15h – 16h (n15)	8
16h – 17h (n16)	8
17h – 18h (n17)	8
18h – 19h (n18)	7
19h – 20h (n19)	7
20h – 21h (n20)	5
21h – 22h (n21)	3

Fonte: O Autor (2024)

Foram inicialmente criados seis turnos de 8 horas mais intervalo para almoço, com

sobreposição nos horários de maior demanda.

Quadro 4: Turnos de trabalho

Número de Funcionários por Turno	Horário de trabalho
x1	08h-12h / 13h-17h
x2	09h-13h / 14h-18h
x3	10h-14h / 15h-19h
x4	11h-15h / 16h-20h
x5	12h-16h / 17h-21h
x6	13h-17h / 18h-22h

Fonte: O Autor (2024)

A função objetivo do modelo é minimizar o número total de funcionários distribuídos entre os seis turnos de trabalho (x1 a x6), que cobrem o funcionamento completo da loja, das 8h às 22h, mantendo a operação eficiente. Ela pode ser descrita pela seguinte equação:

$$\text{Minimizar: } x1 + x2 + x3 + x4 + x5 + x6$$

As restrições garantem que, em cada período do dia, o número mínimo necessário de funcionários esteja presente para atender a demanda de clientes. A tabela inicial define o número mínimo de funcionários necessários por intervalo de uma hora ao longo do dia, variando entre 3 e 8 colaboradores.

As restrições são organizadas de acordo com os turnos que cobrem cada período. Como cada turno cobre uma sobreposição de horas com outros turnos, o número de funcionários presentes em cada hora do dia é calculado como a soma dos funcionários alocados em todos os turnos que cobrem aquele intervalo de tempo.

As restrições podem ser descritas da seguinte forma:

- **R1:** O número de funcionários alocados no primeiro turno (x1) deve ser pelo menos igual ao número mínimo necessário para o período das 8h às 9h (n8). Ou seja:

$$x1 \geq n8$$

- **R2:** A soma dos funcionários dos turnos x_1 e x_2 deve ser pelo menos igual ao número necessário para o período das 9h às 10h (n_9):

$$x_1 + x_2 \geq n_9$$

- **R3:** A soma dos funcionários dos turnos x_1 , x_2 e x_3 deve ser suficiente para cobrir o período das 10h às 11h (n_{10}):

$$x_1 + x_2 + x_3 \geq n_{10}$$

- **R4:** A soma dos funcionários dos turnos x_1 , x_2 , x_3 e x_4 deve ser suficiente para cobrir o período das 11h às 12h (n_{11}):

$$x_1 + x_2 + x_3 + x_4 \geq n_{11}$$

- **R5:** A partir desta restrição começa a se considerar o horário de almoço por turno, sendo assim ela é a soma dos funcionários dos turnos x_2 , x_3 , x_4 e x_5 deve ser suficiente para cobrir o período das 12h às 13h (n_{12}):

$$x_2 + x_3 + x_4 + x_5 \geq n_{12}$$

- **R6:** A soma dos funcionários dos turnos x_1 , x_3 , x_4 , x_5 e x_6 deve ser suficiente para cobrir o período das 13h às 14h (n_{13}):

$$x_1 + x_3 + x_4 + x_5 + x_6 \geq n_{13}$$

- **R7:** A soma dos funcionários dos turnos x_1 , x_2 , x_4 , x_5 e x_6 deve ser suficiente para cobrir o período das 14h às 15h (n_{14}):

$$x_1 + x_2 + x_4 + x_5 + x_6 \geq n_{14}$$

- **R8:** A soma dos funcionários dos turnos x_1 , x_2 , x_3 , x_5 e x_6 deve ser suficiente para cobrir o período das 15h às 16h (n_{15}):

$$x_1 + x_2 + x_3 + x_5 + x_6 \geq n_{15}$$

- **R9:** A soma dos funcionários dos turnos x_1 , x_2 , x_3 , x_4 e x_6 deve ser suficiente para cobrir o período das 16h às 17h (n_{16}):

$$x_1 + x_2 + x_3 + x_4 + x_6 \geq n_{16}$$

- **R10:** A partir desta restrição os primeiros turnos começam a encerrar, sendo assim, a soma dos funcionários dos turnos x_2 , x_3 , x_4 e x_5 deve ser suficiente para cobrir o período das 17h às 18h (n_{17}):

$$x_2 + x_3 + x_4 + x_5 \geq n_{17}$$

- **R11:** A soma dos funcionários dos turnos x_3 , x_4 , x_5 e x_6 deve ser suficiente para

cobrir o período das 18h às 19h (n18):

$$x_3 + x_4 + x_5 + x_6 \geq n_{18}$$

- **R12:** A soma dos funcionários dos turnos x_4 , x_5 e x_6 deve ser suficiente para cobrir o período das 19h às 20h (n19):

$$x_4 + x_5 + x_6 \geq n_{19}$$

- **R13:** A soma dos funcionários dos turnos x_5 e x_6 deve ser suficiente para cobrir o período das 20h às 21h (n20):

$$x_5 + x_6 \geq n_{20}$$

- **R14:** O número de funcionários alocados no último turno (x_6) deve ser pelo menos igual ao número mínimo necessário para o período das 21h às 22h (n21).

$$x_6 \geq n_{21}$$

Cada uma dessas restrições garante que a demanda por funcionários seja atendida em cada período, respeitando os limites mínimos necessários para garantir a operação adequada da loja.

Ao fazer a análise no Solver, foi possível identificar que apenas quatro turnos eram suficientes para atender à demanda de forma eficiente. A tabela abaixo mostra a distribuição dos funcionários ao longo do dia:

Tabela 7: Número de funcionários por turno

Turno	Horário de trabalho	Número de funcionários por turno
1	08h – 17h	3
2	09h – 18h	4
3	10h – 19h	0
4	11h – 20h	0
5	12h – 21h	3
6	13h – 22h	4

Fonte: O Autor (2024)

Dessa forma, foi definido que a loja precisa de, no mínimo, 14 funcionários em operação todos os dias. No entanto, considerando que cada funcionário trabalha 6 dias por semana e tem direito a 1 dia de folga, o número total de funcionários necessário é obtido multiplicando-se o número mínimo diário pela proporção de dias a serem cobertos. Aplicando a fórmula:

$$14 * 7/6 = 16,34$$

Assim, chega-se ao número total de 17 funcionários, garantindo que todas as folgas sejam respeitadas e que a loja esteja sempre adequadamente atendida. No quadro abaixo é possível verificar uma possível distribuição de turnos e folgas para os funcionários ao longo da semana.

Quadro 5: Distribuição de turnos e folgas ao longo da semana

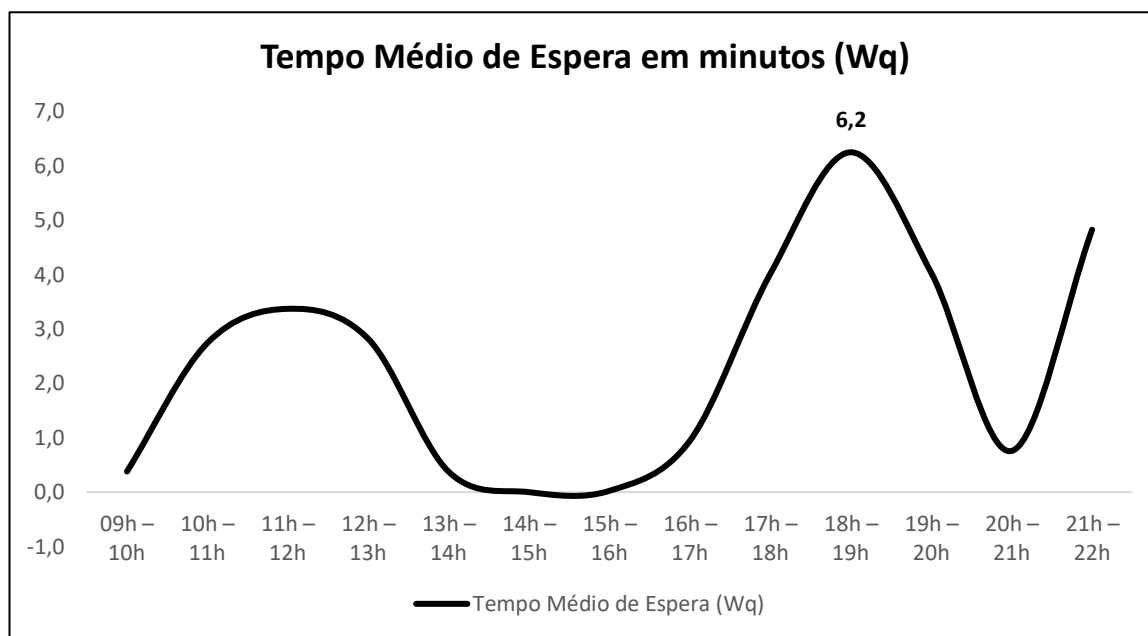
Funcionário	Segunda	Terça	Quarta	Quinta	Sexta	Sábado	Domingo	Folga
1	Folga	08h – 17h	08h – 17h	08h – 17h	08h – 17h	08h – 17h	09h – 18h	Segunda
2	08h – 17h	08h – 17h	08h – 17h	08h – 17h	08h – 17h	Folga	08h – 17h	Sábado
3	08h – 17h	08h – 17h	08h – 17h	Folga	08h – 17h	08h – 17h	08h – 17h	Quinta
4	Folga	09h – 18h	09h – 18h	09h – 18h	09h – 18h	09h – 18h	12h – 21h	Segunda
5	09h – 18h	09h – 18h	09h – 18h	09h – 18h	09h – 18h	Folga	09h – 18h	Sábado
6	09h – 18h	09h – 18h	Folga	09h – 18h	09h – 18h	09h – 18h	09h – 18h	Quarta
7	09h – 18h	09h – 18h	09h – 18h	09h – 18h	Folga	09h – 18h	09h – 18h	Sexta
8	12h – 21h	12h – 21h	12h – 21h	12h – 21h	12h – 21h	12h – 21h	Folga	Domingo
9	12h – 21h	12h – 21h	12h – 21h	12h – 21h	12h – 21h	Folga	12h – 21h	Sábado
10	12h – 21h	12h – 21h	12h – 21h	Folga	12h – 21h	12h – 21h	12h – 21h	Quinta
11	13h – 22h	13h – 22h	13h – 22h	13h – 22h	13h – 22h	13h – 22h	Folga	Domingo
12	13h – 22h	13h – 22h	13h – 22h	Folga	13h – 22h	12h – 21h	13h – 22h	Quinta
13	13h – 22h	13h – 22h	Folga	13h – 22h	13h – 22h	13h – 22h	13h – 22h	Quarta
14	13h – 22h	13h – 22h	13h – 22h	13h – 22h	Folga	13h – 22h	13h – 22h	Sexta
15	09h – 18h	Folga	09h – 18h	12h – 21h	09h – 18h	09h – 18h	12h – 21h	Terça
16	12h – 21h	12h – 21h	12h – 21h	12h – 21h	12h – 21h	12h – 21h	Folga	Domingo
17	08h – 17h	08h – 17h	Folga	08h – 17h	08h – 17h	08h – 17h	08h – 17h	Quarta

Fonte: O Autor (2024)

Nos gráficos a seguir, que ilustram o desempenho do modelo M/M/k em termos de tempo médio de espera (W_q) e utilização dos servidores (ρ), observa-se que, com o número mínimo de 14 funcionários — que conforme a distribuição de turnos do Quadro 5, acontece nas quartas, quintas, sábados e domingos —, o tempo máximo de espera do cliente é de 6,2 minutos. Esse valor está significativamente abaixo da média de 10 minutos estabelecida

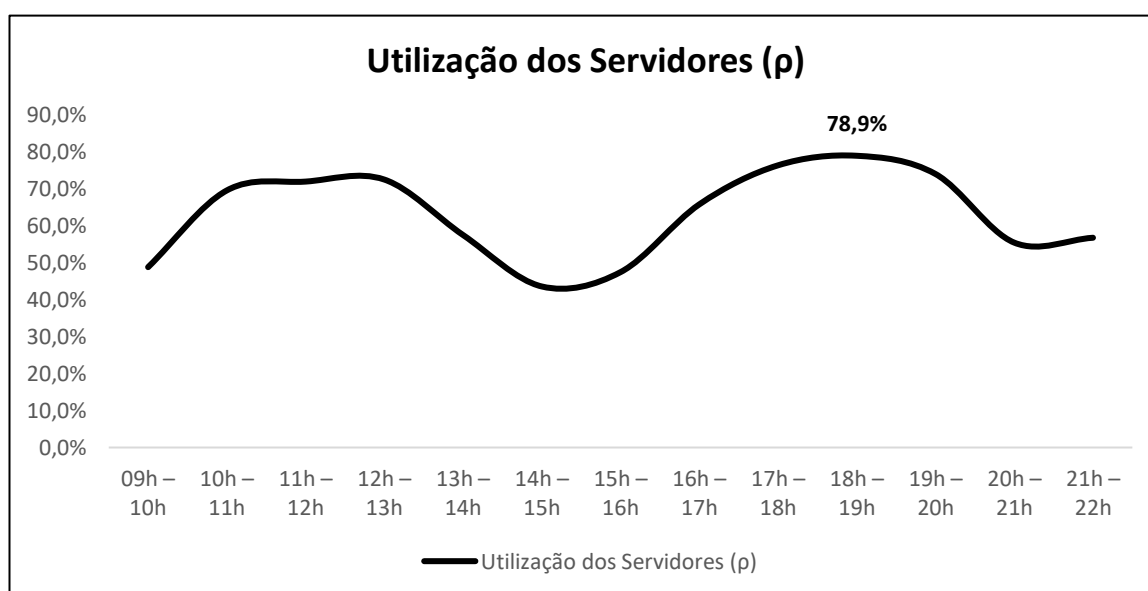
pela empresa. Além disso, a utilização máxima dos servidores, que com os 8 funcionários necessários para o funcionamento era de 82,9%, foi reduzida para 78,9%, sinalizando uma redução da sobrecarga de trabalho, mas ainda sem gerar uma grande ociosidade.

Gráfico 4: Tempo Médio de Espera após elaboração dos turnos (quartas, sextas, sábados e domingos)



Fonte: O Autor (2024)

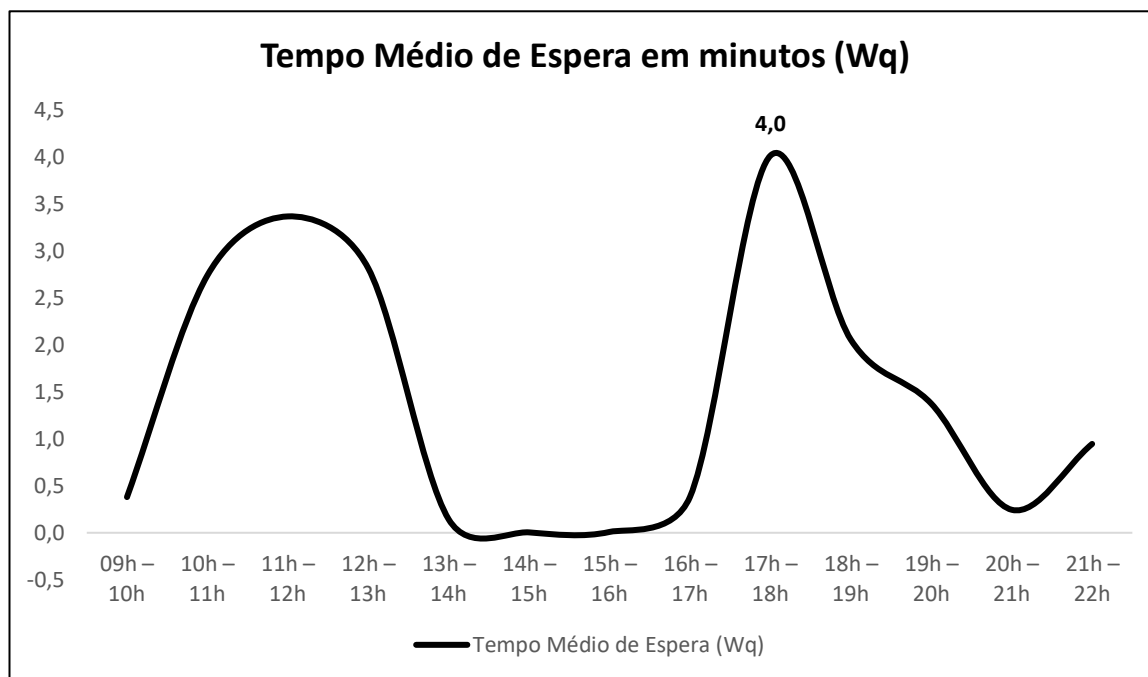
Gráfico 5: Utilização dos Servidores após elaboração dos turnos (quartas, sextas, sábados e domingos)



Fonte: O Autor (2024)

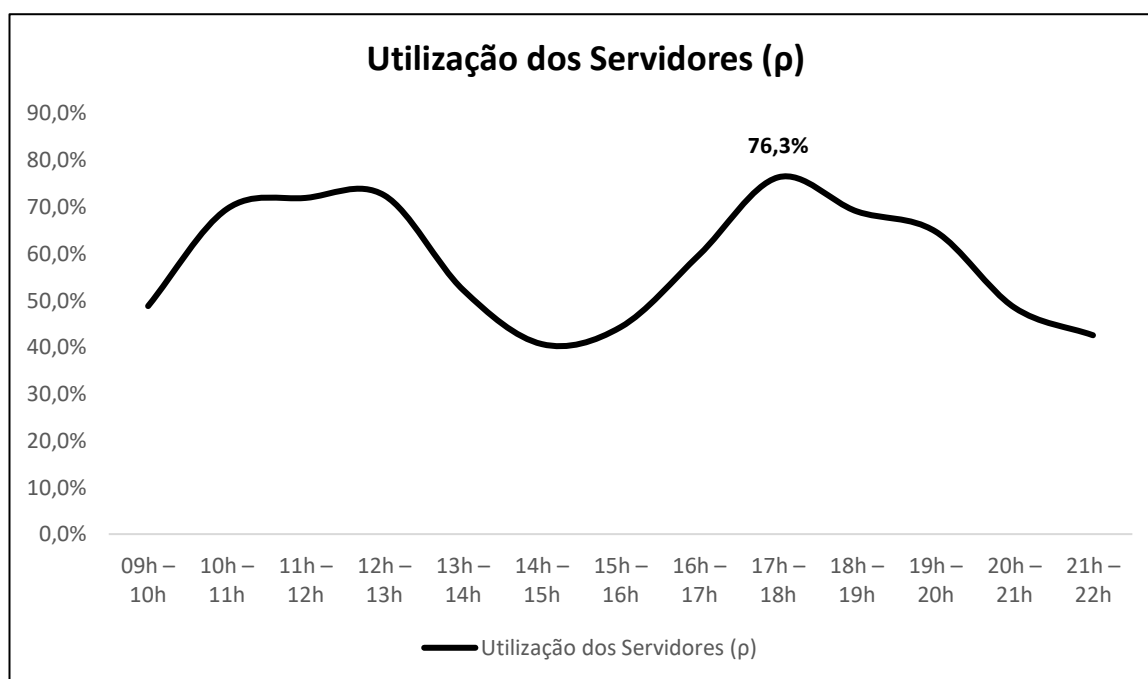
Segundo a distribuição de turnos e folgas ao longo da semana do Quadro 5, para as segundas e terças o tempo máximo de espera do cliente é de 4 minutos e a utilização máxima dos servidores é de 76,3%, como pode ser verificado nos gráficos abaixo.

Gráfico 6: Tempo Médio de Espera após elaboração dos turnos (segundas)



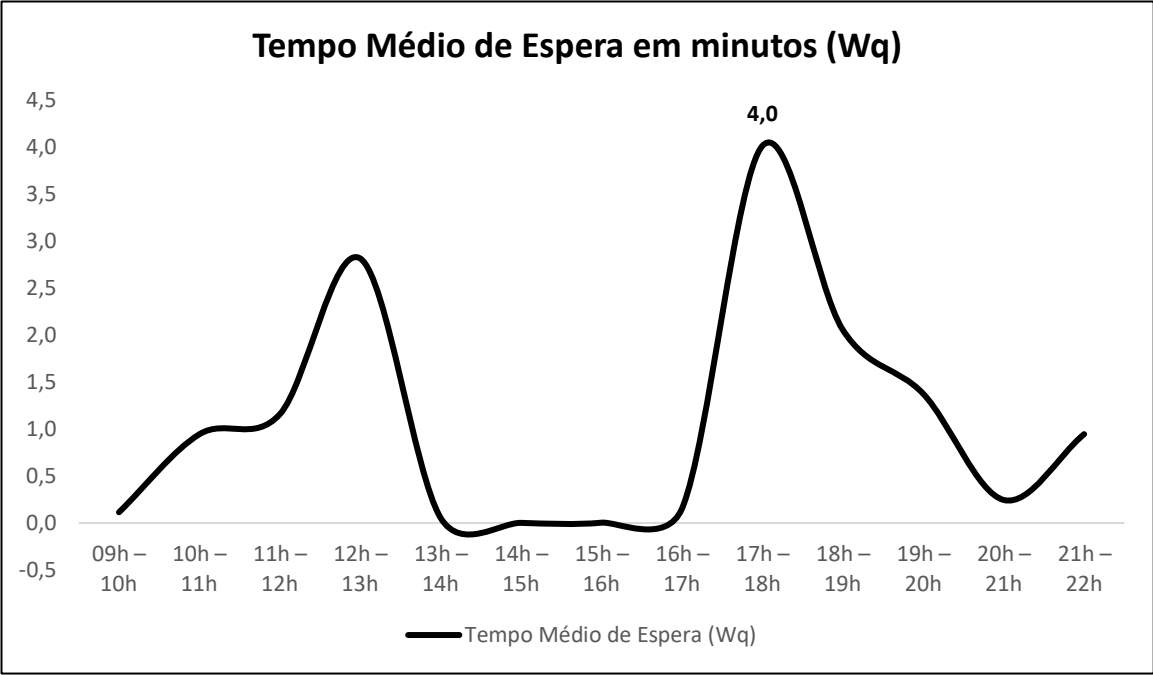
Fonte: O Autor (2024)

Gráfico 7: Utilização dos Servidores após elaboração dos turnos (segundas)



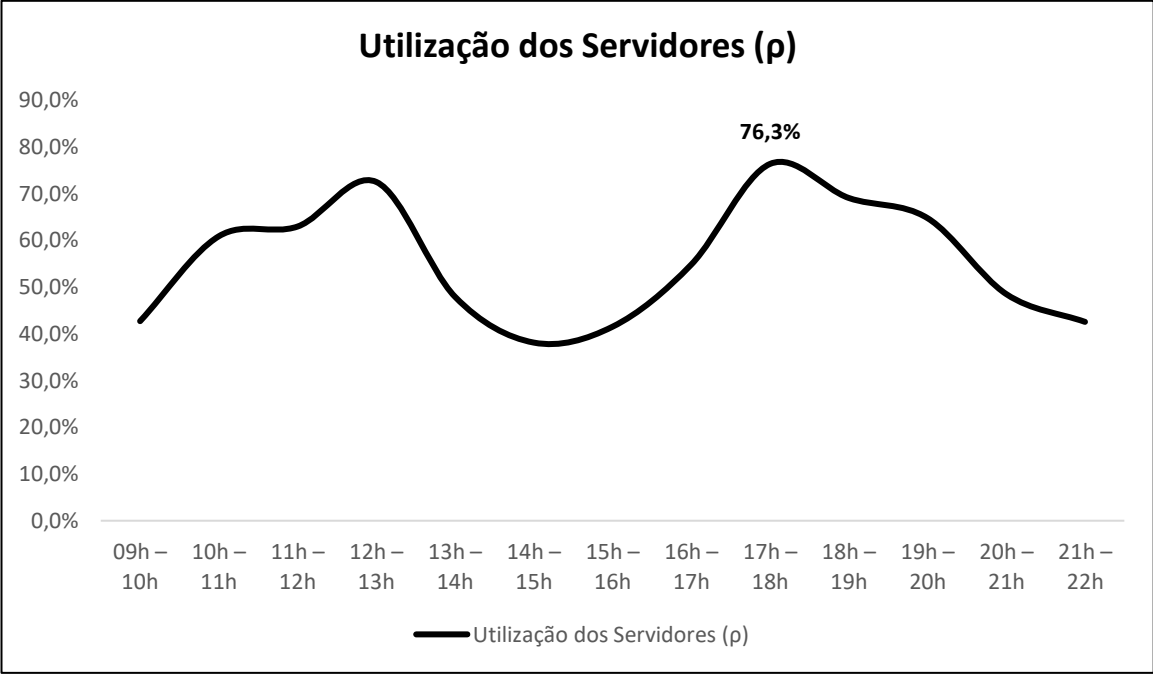
Fonte: O Autor (2024)

Gráfico 8: Tempo Médio de Espera após elaboração dos turnos (terças)



Fonte: O Autor (2024)

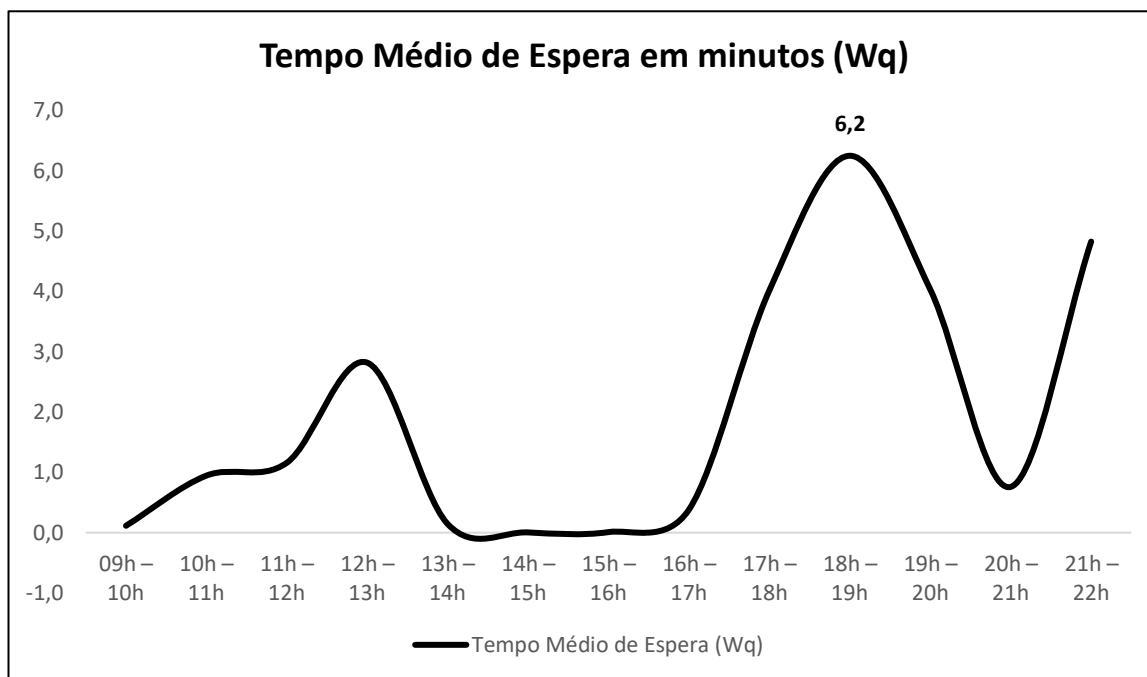
Gráfico 9: Utilização dos Servidores após elaboração dos turnos (terças)



Fonte: O Autor (2024)

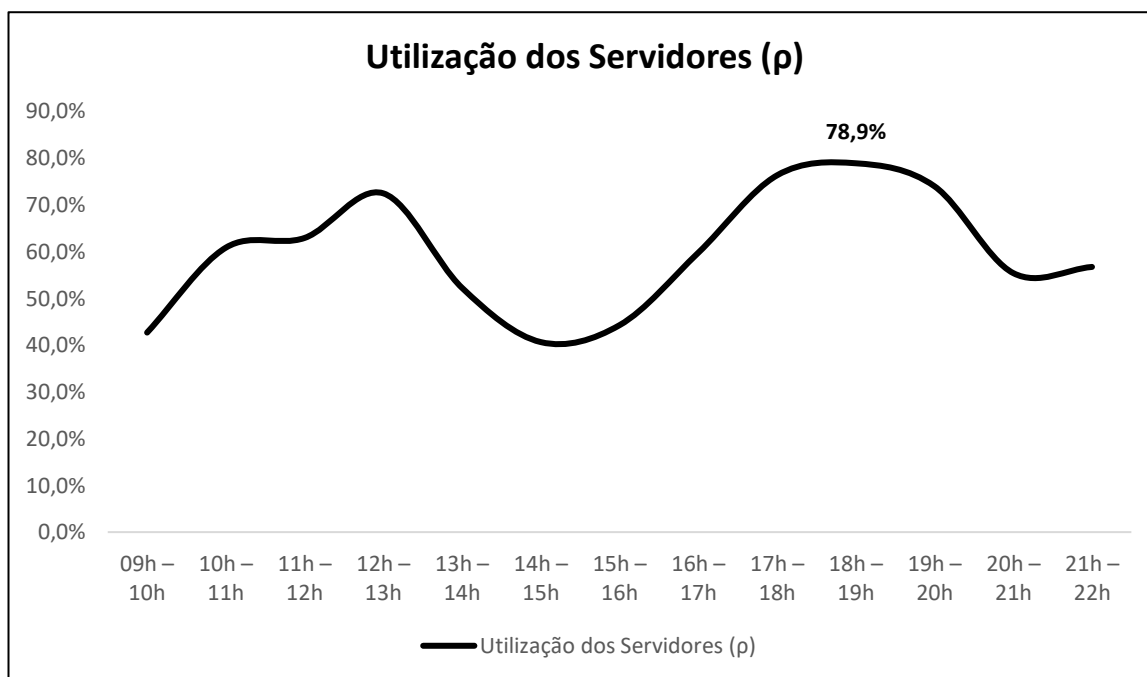
Nas sextas-feiras, mesmo com mais servidores no expediente do dia, tanto o tempo máximo de espera dos clientes quanto a utilização máxima dos servidores são iguais aos das quartas, quintas, sábados e domingos, ou seja, 6,2 minutos e 78,9%, respectivamente.

Gráfico 10: Tempo Médio de Espera após elaboração dos turnos (sextas)



Fonte: O Autor (2024)

Gráfico 11: Utilização dos Servidores após elaboração dos turnos (sextas)



Fonte: O Autor (2024)

4. CONSIDERAÇÕES FINAIS

Este trabalho teve como objetivo aplicar tanto a Teoria das Filas quanto a Programação Linear para otimizar a alocação de funcionários em uma loja de telefonia. Através de uma análise detalhada dos fluxos de clientes e da demanda de atendimento ao longo do dia, foi possível determinar o número ideal de atendentes para garantir um serviço eficiente, minimizando o tempo de espera dos clientes e maximizando a utilização dos recursos humanos.

A aplicação do modelo M/M/k provou ser eficiente ao equilibrar a oferta de atendimento com a demanda, especialmente nos horários de pico, quando a quantidade de funcionários foi ajustada para evitar sobrecargas. Esse modelo, juntamente com a Programação Linear, auxiliou na organização dos turnos de trabalho, respeitando as regulamentações trabalhistas e assegurando uma rotatividade adequada dos funcionários, garantindo que a operação da loja continue eficiente ao longo de todos os dias da semana.

A principal contribuição da Programação Linear neste contexto foi a definição de uma alocação otimizada de recursos, levando em consideração as diferentes restrições, como o número mínimo de funcionários por turno e horários de almoço. Ao utilizar o Solver para resolver o problema de alocação de turnos, foi possível encontrar a melhor solução para minimizar o número de funcionários necessários, sem comprometer a qualidade do atendimento.

Embora o modelo M/M/k e a Programação Linear forneçam uma base sólida para otimizar a alocação de funcionários, algumas limitações foram identificadas. No caso da Teoria das Filas, a suposição de que os tempos de serviço seguem uma distribuição exponencial pode não representar com precisão a variabilidade real do atendimento. Já a Programação Linear, embora eficaz na otimização de turnos, depende fortemente de dados históricos e pode não capturar eventos excepcionais ou variações sazonais.

Um dos principais resultados obtidos com a pesquisa foi a constatação de que a loja, que atualmente conta com 18 atendentes, poderia operar de maneira eficaz com 17 funcionários. Esse dimensionamento reduzido é suficiente para atender à demanda, principalmente nos horários de maior movimento, mantendo a qualidade do atendimento ao cliente. Além disso, foi possível otimizar o fluxo de atendimento, reduzindo o tempo médio de espera e melhorando a utilização dos atendentes.

Conclui-se, portanto, que a combinação da Teoria das Filas e da Programação Linear

é uma ferramenta poderosa para a otimização de processos em ambientes de atendimento ao cliente. Ela não apenas melhora a experiência do cliente, ao reduzir o tempo de espera, como também proporciona uma gestão mais eficiente dos recursos humanos, gerando economia e aumentando a produtividade.

Para trabalhos futuros, recomenda-se explorar outros modelos de filas em ambientes com características operacionais distintas, como aqueles com maior variabilidade no tempo de atendimento. Também seria interessante ampliar o uso da Programação Linear para considerar sazonalidade e eventos excepcionais, ajustando a alocação de funcionários de acordo com as flutuações no fluxo de clientes ao longo do ano.

REFERÊNCIAS

- ARENALES, M.; ARMENTANO, V.; MORABITO, R.; DA FONSECA, C. M. M.. **Introdução à Pesquisa Operacional**. São Paulo: Editora da Unicamp, 2007
- BAHADORI, M.; MOHAMMADNEJHAD, S. M.; RAVANGARD, R.; TEYMOURZADEH, E. Using Queuing Theory and Simulation Model to Optimize Hospital Pharmacy Performance. **Iranian Red Crescent Medical Journal** v. 16.3, E16807, 2014
- BELFIORE, P.; FAVERO, L. P. **Estatística Aplicada**. São Paulo: Saraiva, 2013
- BORDOLOI, S.; FITZSIMMONS, J.; FITZSIMMONS, M. **Service Management: Operations, Strategy, Information Technology**. 10 ed. New York: McGraw-Hill, 2023.
- CHEN, X.; LI Z.; LI, L.; SHI, Q. A Traffic Breakdown Model Based on Queueing Theory. **Networks and Spatial Economics**. v. 14.3-4, p. 485-504, 2014
- FARIAS, A. C.; SIVERIS J. A.; RUPENTHAL, I. L. Teoria das filas aplicada à agenda de atividades de uma empresa de telecom. **Revista CIATEC-UPF**. v. 13.2, p. 169-81, 2021.
- GIL, A.C. **Métodos e técnicas de pesquisa**. São Paulo: Atlas, 1988
- GIL, A.C. **Métodos e técnicas da pesquisa social**. 6ed. São Paulo: Atlas, 2008.
- GOLDBARG, M. C.; LUNA, H. P. L. **Otimização Combinatória e Programação Linear**. Rio de Janeiro: Elsevier, 2005.
- GROSS, D.; HARRIS, C. M. **Fundamentals of queueing theory**. Wiley, 1985
- HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à Pesquisa Operacional**. 8ed. Porto Alegre: AMGH, 2006
- LAKATOS, E. M.; MARCONI, M. A. **Metodologia científica**. Editora Atlas, 1991
- LANTZ, B.; ROSÉN, P. Measuring Effective Capacity in an Emergency Department. **Journal of Health Organization and Management** v. 30.1, p. 73-84, 2016
- LIMA, V. C.; ANHOLON, R; RAMPASSO, I. S.; QUELHAS, O. L. G. Aplicação Da Teoria Das Filas Em Serviços Bancários. **Revista Produção Online**. v. 16.1, p.210, 2016
- MARINS, FERNANDO AUGUSTO SILVA. **Introdução a Pesquisa Operacional**. São Paulo: Cultura Acadêmica: Universidade Estadual Paulista, Pró - Reitoria de Graduação, 2011.
- MORABITO, R.; LIMA, F. C. R. Um modelo para analisar o problema de filas em caixas de supermercados: um estudo de caso. **Pesquisa Operacional**. v. 20.1, p.59-71, 2000
- SILVA, E. L.; MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**.

Universidade Federal de Santa Catarina, 2005

SILVA, T. A.; MALDONADO M. Avaliação do uso de leitos de UTI de covid-19 em Uberaba por meio da teoria das filas. **REMAT: Revista Eletrônica da Matemática**, Bento Gonçalves, RS, v. 7, n. 2, p. e3004, 2021

TAHA, H. A. **Pesquisa Operacional**. 8ed. São Paulo: Pearson Prentice Hall, 2008

WEISS, E.N.; TUCKER, C. Queue management: Elimination, expectation, and enhancement. **Business Horizons**, v. 61, n. 5, p. 671-678, 2018.

WILER, J. L.; BOLANDIFAR, E.; GRIFFEY, R. T.; POIRIER, R. F.; OLSEN T.; FU, R. An Emergency Department Patient Flow Model Based on Queueing Theory Principles. **Academic Emergency Medicine**. v. 20.9, p. 939-46, 2013

YI, W., & WANG, S. (2017). Mixed-Integer Linear Programming on Work-Rest Schedule Design for Construction Sites in Hot Weather. **Computer-Aided Civil and Infrastructure Engineering**. v. 32, p. 429-439, 2017.

ZANARDI, R.; VIALI L. Teoria de filas e o dimensionamento de servidores: um estudo de caso. **Exacta**. v. 18.3, p.489-501, 2020

APÊNDICE A – CÓDIGO TESTE DE KOLMOGOROV-SMIRNOV

```
1 import numpy as np
2 from scipy import stats
3
4 dados = np.array([10.23728814, 14.59322034, 15.09322034, 17.38135593,
5 | 17.22881356, 18.27966102, 19.90677966, 19.71186441,
6 | 18.30508475, 16.55932203, 15.52542373, 11.62711864,
7 | 4.923728814, 5.10169492])
8
9 media_estimada = np.mean(dados)
10
11 def ecdf(data):
12 | sorted_data = np.sort(data)
13 | cdf = np.arange(1, len(data) + 1) / len(data)
14 | return sorted_data, cdf
15
16 def poisson_cdf(data, mu):
17 | return stats.poisson.cdf(data, mu)
18
19 sorted_data, ecdf_data = ecdf(dados)
20 poisson_cdf_data = poisson_cdf(sorted_data, media_estimada)
21
22 ks_statistic, p_value = stats.kstest(dados, 'poisson', args=(media_estimada,))
23
24 print(f"Média estimada (lambda): {media_estimada:.4f}")
25 print(f"Estatística KS: {ks_statistic:.4f}")
26 print(f"Valor-p: {p_value:.4f}")
27
28 if p_value < 0.05:
29 | print("A hipótese nula (distribuição de Poisson) é rejeitada. Os dados não seguem uma distribuição de Poisson.")
30 else:
31 | print("Não rejeitamos a hipótese nula. Os dados podem seguir uma distribuição de Poisson.")
```

Média estimada (lambda): 14.6053

Estatística KS: 0.2513

Valor-p: 0.2883

Não rejeitamos a hipótese nula. Os dados podem seguir uma distribuição de Poisson.