



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Descomplicar a Logística Hospitalar: Introdução da Inteligência Artificial no fluxo logístico de consumíveis clínicos entre o Centro Logístico e o Bloco Central Indiferenciado da CUF Tejo

Rita Alexandra Valentim Picado

Mestrado em Gestão de Serviços de Saúde

Orientador: Prof. João Carlos Rosmaninho de Menezes

Setembro de 2024

Departamento de Marketing, Operações e Gestão Geral

Descomplicar a Logística Hospitalar: Introdução da Inteligência Artificial no fluxo logístico de consumíveis clínicos entre o Centro Logístico e o Bloco Central Indiferenciado da CUF Tejo

Rita Alexandra Valentim Picado

Mestrado em Gestão de Serviços de Saúde

Orientador: Prof. João Carlos Rosmaninho de Menezes

Setembro de 2024

Agradecimentos

Um sincero agradecimento aos meus pais que tornaram a oportunidade de ser aluna deste Mestrado numa Instituição tão singular como é o ISCTE, na realidade que vivo hoje, ao entregar este trabalho que tanto me marcou pela exigência emocional aquando da sua elaboração. Ao meu irmão que tornou impossíveis em desafios e aprendizagens e aos meus avós que tantos dias abdiquei de ver mas nunca deixam de estar perto por tanto de mim serem.

Sem esquecer a ajuda fornecida ao longo dos meses de estágio académico no Hospital CUF Tejo, que se tornou no elemento impulsionador do tema escolhido e cujos profissionais se demonstraram prestáveis na partilha de informações e dados.

Por último, um obrigada pelas horas dadas à redação do trabalho ao meu orientador, Prof. João Carlos Rosmaninho de Menezes.

Abstract

Digital transition has been identified as a political priority by the executive body of the European Union. The European Commission has set targets and created supports aimed at promoting the digital transformation of the Member States.

The application of Artificial Intelligence, Machine Learning and predictive data analysis technologies serves as an aid in preparing healthcare service providers to respond to times of high demand from patients. Understanding the utility of its application, supported by algorithms and historical data, it becomes possible to anticipate the need for clinical consumables. Following that and in order to clarify the impact of analyzing historical consumption and supply data of clinical consumables related to BCI, I suggest applying Machine Learning tools to the logistics management of CUF Tejo Unit.

The work will follow an exploratory methodology, integrating qualitative techniques such as informal interviews, with quantitative data analysis where Machine Learning algorithms will be applied to data previously exported from three Excel tables to a Jupyter Notebook, using Python, a programming language.

Expected outcomes include optimized logistical processes at CUF Tejo Unit. To adjust ordered quantities and the maximum level associated with the MRP of each material, it is suggested that any materials exceeding average order quantities should double their maximum stock levels. This new variable will support management in responding to potential needs in the upcoming months. With this approach it may be possible to enhance the efficiency of logistical operations and improve the quality of services provided.

KeyWords: Artificial Intelligence, Healthcare services, Historical data, Logistics management, Machine Learning

JEL Classification: C63, I18

Resumo

A importância da transição digital é de tal ordem que foi identificada como prioridade pelo órgão executivo da União Europeia, a Comissão Europeia, que fixou metas com vista à promoção da transformação digital dos Estados-Membros.

A aplicação de Inteligência Artificial, Machine Learning e das tecnologias de análise preditiva de dados auxilia a preparação das Unidades prestadoras de serviços de saúde. Ao entender a utilidade que a sua aplicação, com suporte em algoritmos e dados históricos, torna-se possível antecipar a necessidade de consumíveis clínicos. Na sequência daquela afirmação, de modo a concluir acerca do impacto que tem a análise de dados históricos de consumo e fornecimento de consumíveis clínicos afetos ao BCI, sugiro a aplicação de ferramentas de Machine Learning na gestão logística da Unidade CUF Tejo.

O trabalho segue uma metodologia exploratória, integrando técnicas qualitativas, como entrevistas informais e a análise de dados quantitativos, previamente exportados de três tabelas do Excel para um Jupyter Notebook, utilizando a linguagem de programação Python.

Finalizada a análise dos dados espera-se que os resultados permitam fazer ilações quanto à otimização dos processos logísticos da Unidade CUF Tejo. De modo a ajustar as quantidades encomendadas e o nível máximo associado ao nível MRP de cada material, sugere-se que qualquer material cuja quantidade média de encomendas seja superior a um valor de referência, deva sofrer um aumento de duas vezes o seu nível máximo MRP, de modo a diminuir o número de vezes que a sua reposição ocorre. A variável criada, de apoio à gestão, dará resposta a hipotéticas necessidades nos meses vindouros.

Palavras-chave: Dados históricos, Gestão logística, Inteligência Artificial, Machine Learning, Serviços de saúde, Transição digital

Classificação JEL: C63, I18

Índice

1	Introdução.....	1
1.1	Motivação.....	3
1.2	Objetivos	4
1.3	Estrutura.....	5
2	Revisão de Literatura	6
2.1	Gestão da Cadeia de Abastecimento	6
2.2	Logística no Bloco Operatório	7
2.3	Inteligência Artificial na Gestão da Cadeia de Abastecimento.....	8
2.4	Machine Learning na GCA.....	9
2.4.1	<i>Pipeline</i> da Ciência de Dados	12
2.5	Desafios e benefícios da IA na GCA	13
2.6	Modelos de previsão.....	13
3	Metodologia de investigação.....	15
3.1	Técnicas e instrumentos de recolha de dados.....	15
3.2	Metodologia da Análise dos Dados	15
4	Logística na CUF Tejo.....	17
5	Dados.....	19
5.1	Recolha.....	19
5.2	Compreensão e preparação.....	19
5.3	Análise.....	20
6	Resultados	31
7	Discussão de resultados	35
8	Conclusão.....	37
8.1	Implicações do trabalho para a gestão	38
8.2	Limitações	38
8.3	Sugestões	39
8.3.1	Sugestões para investigação	40
8.3.2	Sugestões para o Hospital.....	40
9	Referências.....	43

Índice de Figuras

Figura 1. Estrutura CUF	2
Figura 2. Etapas da entrada de uma encomenda	3
Figura 3. Quadro Concetual	6
Figura 4. Pipeline do processo de aprendizagem supervisionada	9
Figura 5. Publicações sobre IA de estudantes portugueses	10
Figura 6. Utilização de IA na Europa.....	11
Figura 7. Cross Industry Standard Process for Data Mining (Shearer, 2000, p. 14)	11
Figura 8. Hierarquia da Inteligência Artificial.....	16
Figura 9. Gestão dos consumíveis clínicos	17
Figura 10. Descrição das variáveis referentes aos dados a analisar	20
Figura 11. Índices de rutura	21
Figura 12. Evolução índices de rutura	23
Figura 13. Correlação rutura entre variáveis	24
Figura 14. Scatterplots rutura entre variáveis	25
Figura 15. Distribuição estatística das variáveis	26
Figura 16. Relação Nível Mínimo MRP e Quantidade Entrada no BCI	27
Figura 17. Quantidade entrada no armazém do BCI	27
Figura 18. Distribuição dos materiais por classe ABC	28
Figura 19. Top 5 materiais de cada classe	29
Figura 20. Nº de presenças mensais de determinado cod_sap numa encomenda	32
Figura 21. Variável binária	33

Glossário de siglas

AA	Armazém Avançado
BO	Bloco Operatório
BCI	Bloco Central Indiferenciado
CL	Centro Logístico
GCA	Gestão da Cadeia de Abastecimento
IA	Inteligência Artificial
LH	Logística Hospitalar
ML	Machine Learning
MRP	Material Requirements Planning

1 Introdução

Às exigências do mercado, estão implícitas as exigências feitas pelos clientes que priorizam, cada vez mais, um atendimento rápido e que lhes seja o mais útil possível. Deste modo, cabe às empresas reter esses clientes, evitando situações indesejadas pelos mesmos, como é exemplo a espera por determinado produto ou serviço num mercado cheio de opções de escolha, até dentro da mesma instituição.

Como consequência à concretização do referido, qualquer Unidade Hospitalar deve promover a criação de valor para o cliente, a diferenciação da oferta, a organização dos cuidados de saúde, a satisfação do corpo clínico e a qualidade dos cuidados prestados.

O Hospital CUF Infante Santo, atual CUF Tejo e primeira Unidade de saúde da rede, inaugurada em 2020, pertence a uma cadeia que está, à data, presente em 15 municípios portugueses, através de 24 Hospitais e Clínicas, que atuam em rede, se complementam na prestação de cuidados e agem de forma indispensável na promoção da saúde em Portugal. A CUF tem consolidado, ao longo dos anos, a sua missão de promover a prestação de serviços de saúde com os mais elevados níveis de qualidade e conhecimento. Esta Unidade, prestadora de serviços de saúde, apresenta-se como um dos projetos mais inovadores e diferenciadores no setor privado da saúde em Portugal. O Hospital promove ainda a investigação clínica e a formação universitária e pós-graduada em saúde.

Com vista a alcançarem um melhor desempenho e o aumento da capacidade competitiva através da gestão de inventário (Chalotra, Vipul., 2013), as organizações têm o dever de promover o desenvolvimento e a implementação de procedimentos de gestão de stocks com vista à redução de desperdícios e custos operacionais (Verhoef e Sloat, 2006).

A criação do Centro Logístico data de 2019. De modo a sustentar a sua reputação como uma das pioneiras do crescimento sustentado no desenvolvimento e implementação de uma excelente gestão do fluxo de materiais de consumo clínico a CUF, para além de desempenhar as suas funções internas através da existência e correta utilização de Armazéns Avançados, centralizou a logística das Unidades detentoras daqueles armazéns com a criação do Centro Logístico, CL CUF. O CL CUF funciona ainda sob alçada de diversas limitações, a identificar e que se refletem não só no clima organizacional como também na conservação das relações a manter com fornecedores externos.

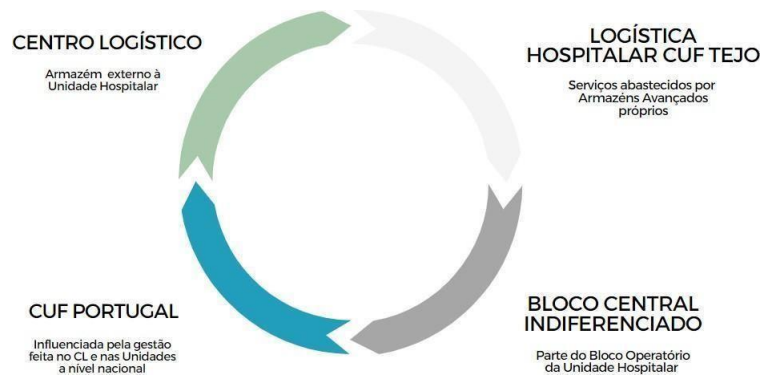


Figura 1. Estrutura CUF

Subjacente à sua implementação estão diversos motivos que a justificam, entre eles:

- A criação de um espaço físico cujo custo por m2 fosse inferior ao das Unidades que já possuíam armazéns centrais internos, passando os custos a ser referentes apenas ao transporte entre o Centro Logístico e a Unidade e à deslocação dos fornecedores;
 - A gestão mais eficaz de materiais por parte das Unidades sem armazém central, por os mesmos serem previamente preparados à unidade pelo CL CUF;
 - A existência de um espaço com staff despreocupado com as relações hospitalares, sendo contratados operadores internos ao CL CUF, especializados apenas na preparação e recepção de mercadorias, tornando o fluxo mais rápido, promovendo a produtividade, a diminuição do índice de erro e a redução do fluxo de fornecedores nas imediações das Unidades Hospitalares;
- De modo a tornar claro o fluxo das entradas subjacentes à colocação de uma encomenda, feita com vista à supressão de uma necessidade no Bloco Central Indiferenciado, BCI, o diagrama abaixo descreve sucintamente cada uma daquelas etapas, separando, por cores as afetas ao CL das afetas à Logística Hospitalar..

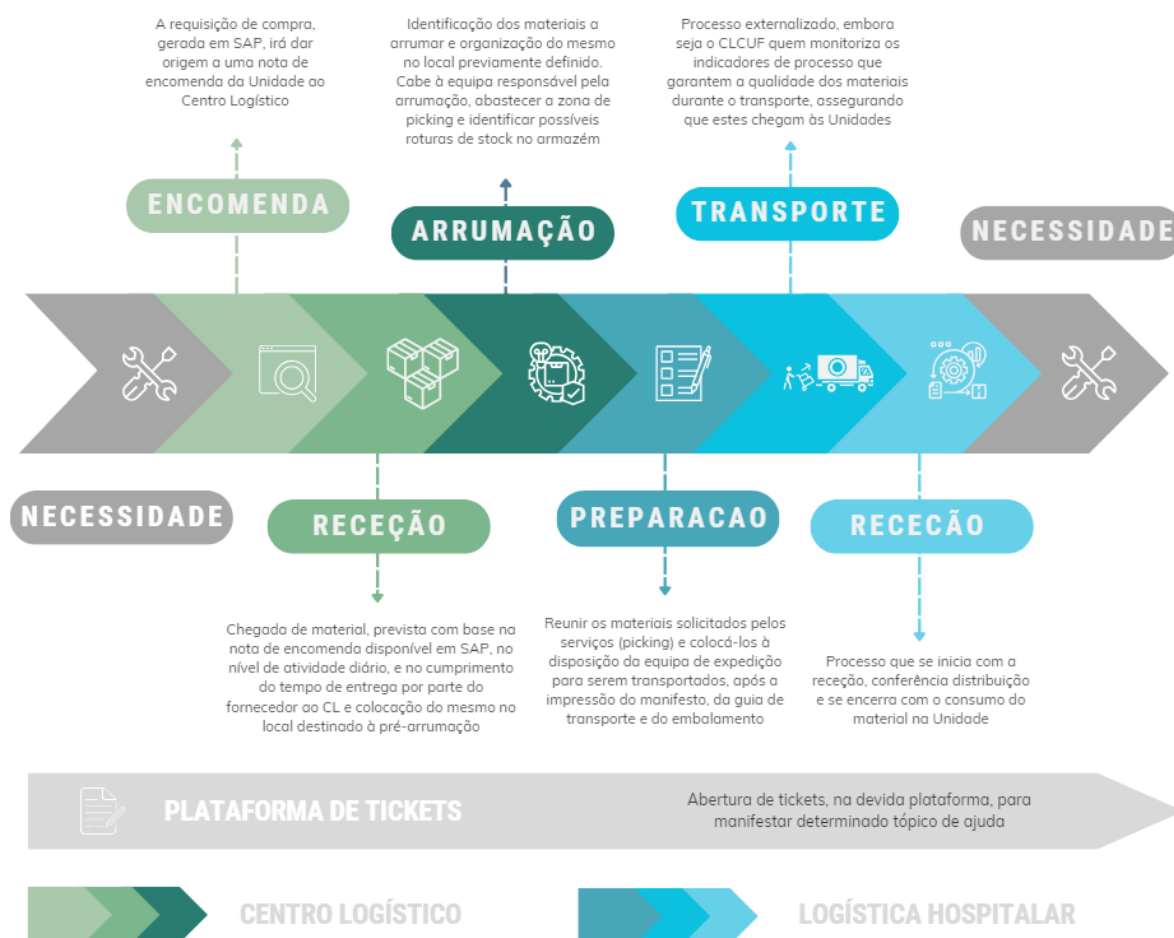


Figura 2. Etapas da entrada de uma encomenda

De modo a esclarecer se a análise de dados históricos referentes aos padrões de consumo e abastecimento de consumíveis clínicos afetos ao BCI contribui para a gestão do abastecimento do CL à Unidade Hospitalar CUF Tejo, sugere-se como título para o trabalho desenvolvido:

- Descomplicar a Logística Hospitalar: Introdução da Inteligência Artificial no fluxo logístico de consumíveis clínicos entre o Centro Logístico e o Bloco Central Indiferenciado da CUF Tejo.

1.1 Motivação

A direção do serviço de compras e logística da CUF Tejo assume como responsabilidades a gestão da previsão daquilo que poderá ser preciso e a organização dos recursos, ou seja, dos materiais de consumo clínico, com a devida antecedência, para que estes possam ser utilizados de forma eficiente.

Aliada àquelas responsabilidades, surge a pressão para a resolução rápida das limitações, com vista à redução dos custos em termos de espaço, quantidade de stocks e mão-de-obra, fatores que correspondem a cerca de 30% a 40% do orçamento das instituições de saúde (Aguilar - Escobar et al, 2015).

Assente numa cultura de melhoria contínua e de forma a consolidar o seu projeto clínico diferenciador, a CUF tem apostado, nos últimos anos, na medição de *outcomes* com base em modelos propostos pelo *The International Consortium for Health Outcomes Measurement* (ICHOM), e em *standards* desenvolvidos pelas equipas clínicas internas. O programa de valor em saúde tem também foco na participação em projetos com redes nacionais e internacionais. São exemplos, a integração no *European Health Data and Evidence Network* (EHDEN) e a participação da CUF no simpósio promovido pelo *Observational Health Data Sciences and Informatics* (OHDSI), realizado em julho de 2023 e cujo objetivo consistiu na organização de dados em saúde.

Apesar dos esforços demonstrados pela Unidade, surge ainda, à data e aliada à gestão logística, a possibilidade de melhoria no que diz respeito à análise de dados através da transformação digital e da possibilidade de manipulação dos mesmos para fornecer informações úteis à gestão hospitalar.

1.2 Objetivos

O objetivo primordial do projeto passa pela compreensão do impacto que poderá ter a análise dos padrões associados ao grupo de consumíveis clínicos afetos ao Bloco Central Indiferenciado, cujo processo logístico implica a sua passagem pelo CL CUF e a existência de regras. Estas regras devem ser internas e esclarecedoras da dinâmica sobre a qual recai o movimento de entrada e saída dos consumíveis clínicos do Catálogo CL CUF, ou seja, do âmbito do Centro Logístico.

Com vista ao cumprimento do objetivo, são apresentados abaixo os objetivos secundários (Fx), que contribuirão para a elaboração faseada do projeto.

F1: Existência de critérios, do âmbito de abastecimento de determinado grupo de materiais de consumo clínico, do Centro Logístico às diversas Unidades Hospitalares.

F2: Existência de dados referentes a variáveis cuja análise permite fazer previsões logísticas acerca de problemas de gestão de stocks.

F3: Informações a recolher, dos dados e pela Logística Hospitalar, de modo a prever necessidades futuras de gestão de stocks e a minimizar ineficiências existentes.

F3.1: Procura de soluções a avaliar e implementar para resolver dificuldades, de modo a mitigar possíveis constrangimentos que surjam aquando da operação logística.

F3.2: Conclusões úteis à Logística Hospitalar com base na análise dos dados recolhidos.

O cumprimento do objetivo principal e dos objetivos secundários propostos acima, será sucedido tendo por base a utilização de uma metodologia que integrará ambas as abordagens qualitativas e quantitativas. A primeira abordagem surge sustentada, por exemplo, em entrevistas, realizadas com o intuito de adquirir informação acerca da forma como se encontra estruturado o departamento logístico da Unidade Hospitalar. Por sua vez, a abordagem quantitativa é realizada através da análise de dados históricos referentes ao consumo e ao abastecimento de consumíveis clínicos, úteis à posterior apresentação de sugestões aquela que é a realidade da logística hospitalar. Esta análise não só pretende auxiliar a gestão logística na supressão das necessidades diárias hospitalares, como servir de exemplo ao leque de possibilidades que estudantes e pesquisadores possuem para promover a mudança em ambientes que a devem adotar, nos seus serviços, o quanto antes, de modo a aumentar a vantagem competitiva quando em comparação com entidades semelhantes.

1.3 Estrutura

O trabalho desenvolvido subdivide-se em capítulos. O segundo capítulo fornece uma visão geral da literatura, que foca a importância do uso de técnicas de Machine Learning no auxílio à gestão, num mercado com tendências visivelmente emergentes, e no qual serão discutidos tópicos como a gestão da cadeia de abastecimento, GCA, a logística hospitalar no Bloco Operatório, a Inteligência Artificial e os seus benefícios na GCA, a aplicação de Machine Learning na GCA e uma abordagem aos modelos de previsão. O foco do capítulo três recai sobre o enquadramento metodológico, sendo apresentadas as técnicas e instrumentos de recolha de dados e a metodologia da análise desses mesmos dados. No quarto capítulo é apresentada, de forma introdutória e com o intuito de clarificar aquela que é a logística na CUF Tejo e a forma como esta funciona, informação referente a estes temas. O capítulo consequente aborda o tema da análise dos dados, cujos resultados são apresentados no capítulo seis e discutidos no sétimo capítulo. O oitavo e último capítulo, considera todo o trabalho e pesquisa efetuados, sendo mencionadas as implicações do trabalho para a gestão, apresentadas as limitações e feitas sugestões para investigações futuras e para a entidade que contribuiu para a realização do trabalho desenvolvido.

2 Revisão de Literatura

O quadro conceitual apresentado na figura 3 fornece um enquadramento dos aspectos a desenvolver no decorrer da elaboração do trabalho. Inicialmente é feita uma introdução a conceitos como a GCA e a presença da Logística Hospitalar no bloco operatório. Mais à frente é descrita a presença da Inteligência Artificial e de Machine Learning na Cadeia de Abastecimento da indústria e são apresentados desafios e benefícios da Inteligência Artificial na Gestão da Cadeia de Abastecimento. O último ponto refere os modelos de previsão.

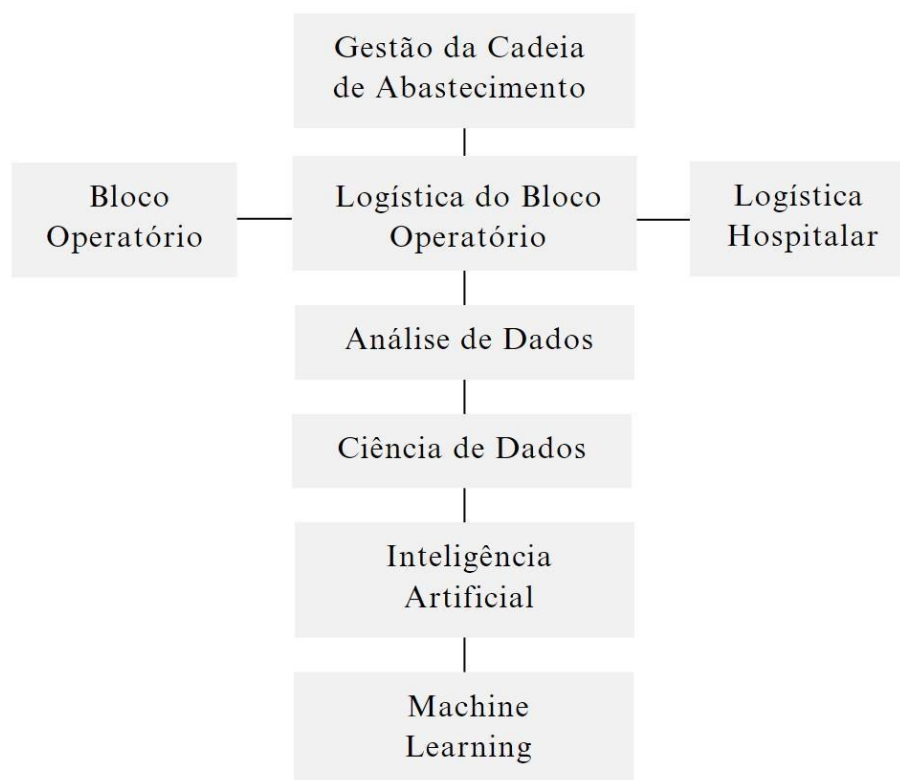


Figura 3. Quadro Conceptual

2.1 Gestão da Cadeia de Abastecimento

Subjacente a uma cadeia de abastecimento adequada encontram-se diversos fatores, entre eles, a redução de custos e riscos (Tarei et al., 2022), a qualidade dos mesmos (Munir et al., 2020) e a flexibilidade (Azadegan et al., 2020). Estes fatores surgem, segundo Koehler, como meios para dar resposta às necessidades dos consumidores de serviços de saúde, que esperam produtos e serviços personalizados; ao aumento do stock de materiais disponíveis; e à redução de desperdícios, promovendo a sustentabilidade (Yousefi e Tosarkani, 2022).

A obtenção de uma cadeia de abastecimento eficaz é considerada, segundo Schuetz e Venkatesh

(2020), como um desafio em diversas empresas, devido à presença de Unidades prestadoras de serviços de saúde abrangidos por diferentes jurisdições (Kwon et al., 2016; Gardas, 2022). Deste modo torna-se perceptível a importância da redução da complexidade e a facilidade associada aos processos da GCA (Ashwani Kumar, Venkatesh Mani, Vranda Jain, Himanshu Gupta, & V.G. Venkatesh., 2023).

A cadeia de abastecimento das unidades de saúde surge associada a uma rede de processos que asseguram a produção, distribuição e o fornecimento atempado de materiais de consumo clínico necessários à prevenção da saúde dos doentes (Beaulieu., Bentahar., 2021) ou à prestação de cuidados aos mesmos.

A disponibilidade de materiais de consumo clínico necessários à prestação de serviços de saúde, conforme menciona Tetteh (2019), possui uma relação direta com o tempo de espera associado à sua cadeia de abastecimento. Por outras palavras, a ocorrência de atrasos no abastecimento e reposição de consumíveis, promove a interrupção do fluxo da cadeia de abastecimento, aumentando o tempo de espera associado aos serviços prestados e impacta de forma indesejada a satisfação dos utentes.

O objetivo da GCA passa por atender às necessidades dos consumidores, otimizando custos associados a stocks e processos. Para garantir que esse objetivo possa ser alcançado é necessário, segundo Monden, Y. (2012), reduzir o *lead time*. A redução dos prazos de entrega por parte dos fornecedores ou do *lead time* consiste numa variável bastante influenciada por fatores externos que levam ao seu difícil controlo. Entre eles surgem questões logísticas, como foram exemplo, conforme exposto no balanço da Flexport, partilhado no seu blogue a 12 de janeiro deste ano, os congestionamentos de tráfego no Mar Vermelho e no Canal do Suez que, de acordo com o jornalista Jorge Nascimento Rodrigues (2024), envolvem 12% do comércio marítimo mundial e 30% do tráfego global de contentores. A acrescentar, tem ainda influência a disponibilidade de matérias primas no local de origem. Considerar estas variáveis e tentar mitigá-las utilizando estratégias de gestão de risco e diversificando os fornecedores, são propostas a ter em conta para reduzir o impacto das oscilações nos prazos de entrega.

2.2 Logística no Bloco Operatório

No Bloco Operatório, local onde profissionais desempenham funções sob pressão, por ser ínfima a possibilidade de erro, é exigido, à equipa cirúrgica, o domínio da cadeia de abastecimento de materiais de consumo clínico, instrumentos cirúrgicos e equipamentos (Lopes, 2012).

Os serviços de saúde, reconhecidos pelo autor Hollnagel (2017) como sistemas complexos, por serem desenvolvidos sob condições e situações esperadas e inesperadas, enquanto mantém o correto fluxo de entrada e saídas de consumíveis clínicos, devem ser geridos tendo em conta o ambiente interno e externo à Unidade onde são prestados, e o impacto que têm na sua eficiência (Machry et al., 2021).

2.3 Inteligência Artificial na Gestão da Cadeia de Abastecimento

Artificial Intelligence e *Machine Learning* são sinónimo de mudança, por promoverem transformações significativas a curto e a longo prazo nos processos logísticos de diversas áreas, como é exemplo a saúde (Habeheh, H., Gohel, S., 2021). A utilização da IA na saúde permite identificar e resolver perturbações no fornecimento de consumíveis clínicos às demais Unidades (SupplyChainBrain., 2024).

A empresa de *media* “*Economist Impact*” entrevistou 3.500 executivos seniores em todo o mundo para a quarta edição da sua pesquisa, intitulada “*Trade in Transition 2024*”, um programa apoiado pela *DP World* (DP World, 2024), dedicado a compreender as principais tendências afetas ao comércio global e às cadeias de abastecimento. No relatório, publicado a 16 de janeiro, cerca de metade das empresas justificam a utilização da IA como meio para melhorar as experiências dos consumidores, enquanto outras 35% utilizam esta tecnologia para prever as exigências feitas pelos mesmos. Dados demonstram ainda que mais de um terço dos participantes da pesquisa (34%) afirmam que o uso da IA ajudou a sua empresa a reduzir o custo geral das operações comerciais. Outros 32% afirmam que a IA permitiu que a sua organização reduzisse as interrupções na cadeia de abastecimento. O apoio prestado à publicação do relatório justifica-se pela opinião defendida por Magrietha Mallinson (DP World, 2024), “*global vertical leader*” de Saúde na DP World, de que numa altura em que a imprevisibilidade surge como “a nova constante”, torna-se imperativo utilizar tecnologias avançadas como a IA, nomeadamente quando os métodos de produção se tornam cada vez mais avançados e no momento no qual a indústria deve inovar todas as fases da cadeia de abastecimento.

A ocorrência da IV Revolução Industrial, associada às tecnologias digitais, fomentou a adoção, por parte das empresas de tecnologias como a AI, blockchain e IoT (Rajput e Singh, 2019; Egger e Masood, 2020). Esta adoção promoveu não só o crescente desenvolvimento de aplicações baseadas em IA para cadeias de abastecimento (Riahi et al., 2021), mas também a exploração do impacto daquela tecnologia na GCA (Saberri et al., 2019).

À data, a nível mundial e de modo que a sua utilização não seja feita de forma desregulada, a plataforma *Global Health AI* surge com o intuito de promover a utilização responsável da inteligência artificial na saúde (Cátia Mateus et al, 2023). Deste modo, tornou-se ainda mais apelativa e segura a aplicação de IA na realização de recomendações, para automatizar processos e melhorar a produtividade das Unidades prestadoras de serviços de saúde (Molina, Arturo. 2023).

2.4 Machine Learning na GCA

O desenvolvimento de algoritmos, dotados da capacidade de aprendizagem (Jan A Hazelzet., 2009), através do uso de técnicas de *Machine Learning* auxilia a gestão, num mercado com tendências emergentes. Subjacente aquela ajuda está a aplicação de métodos de recolha, processamento e organização de dados de diferentes fontes, a sua análise e a identificação de padrões e tendências que auxiliam as organizações a identificar oportunidades de melhoria.

A aplicação de ML tem como objetivo identificar (ou aprender) uma função $f: X \rightarrow Y$ que indica o domínio de entrada X (dados) para o domínio de saída Y (previsões). Tendo em consideração o tipo de conjuntos de dados utilizados para teste, os algoritmos de ML podem ser classificados em três categorias, a aprendizagem supervisionada, a aprendizagem não supervisionada e a aprendizagem por reforço. Na aprendizagem supervisionada, categoria sob a qual será debruçada, numa fase inicial, a atenção e cujo processo se apresenta na figura 4, os sistemas utilizam conjuntos de dados rotulados $(x, y) \in X \times Y$, onde x representa um dado e y representa a previsão correspondente a x (Silva, G., Pereira, L., Carvalho, J., Vinhas da Silva, R., & Simões, A., 2024).

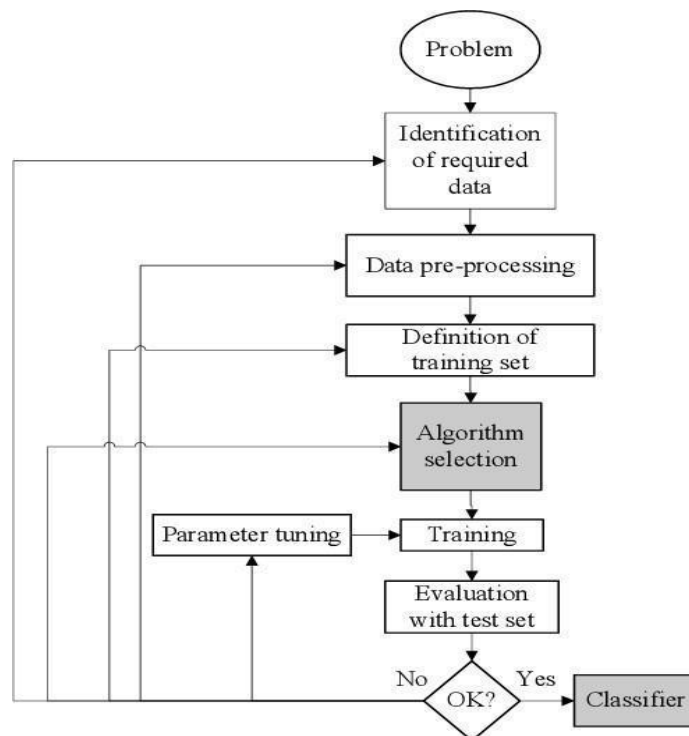


Figura 4. Pipeline do processo de aprendizagem supervisionada

As técnicas de ML identificam e extraem automaticamente padrões existentes entre variáveis através da análise de grandes conjuntos de dados (Tirkolaee, E. B. et al, 2021) . À data, tornam-se, por isso, raras as vezes em que a extração de padrões ou informações de um grande volume de dados e a sua interpretação são possíveis através da utilização de técnicas analíticas tradicionais (A. Dey, 2016). O crescimento da capacidade de processamento e da disponibilidade de dados possibilitou o aumento do investimento em Portugal, através da aplicação de técnicas de ML numa imensa variedade de indústrias. O gráfico abaixo representa o interesse dos estudantes em contribuir, futuramente, para aquele crescimento, algo visível através da observação do aumento do número de publicações, entre 2000 e 2023, sobre o tópico nas principais instituições universitárias de Portugal.

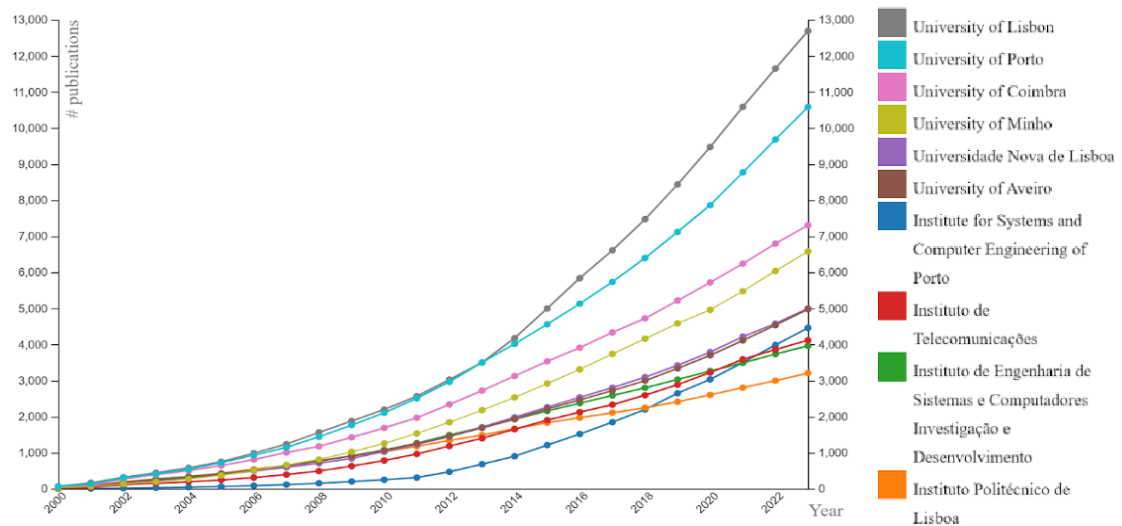


Figura 5. Publicações sobre IA de estudantes portugueses

O relatório “Ambições da UE para a inteligência artificial” redigido por Jörg Genner, Ștefan-Rhklęgas (2024), indica que 35% das empresas portuguesas adotaram tecnologias de IA em 2023, face aos 28% de 2022. Aquele relatório realça ainda que a utilização da Inteligência Artificial e de outras tecnologias digitais está mais concentrada nas empresas de maior dimensão (45% em comparação a apenas 32% das pequenas e médias empresas), conforme se verifica na figura abaixo, referente ao ano de 2021.

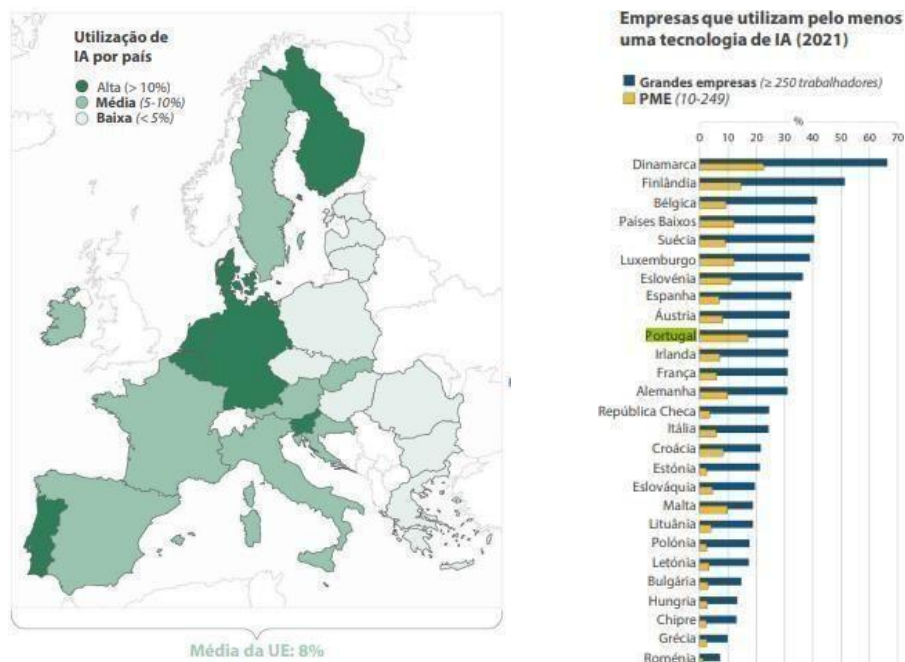


Figura 6. Utilização de IA na Europa

2.4.1 Pipeline da Ciência de Dados

A ciência de dados subdivide a análise passível de fazer aos dados em etapas, organizadas num *pipeline*, onde, imaginando-se uma linha de etapas, conforme demonstra a figura 7 a análise aos dados deve respeitar uma sequência lógica, que começa na aquisição dos dados e passa pelo armazenamento, limpeza, transformação, treino e eventual modelação de algoritmos de ML (Gamma et al., 1993).

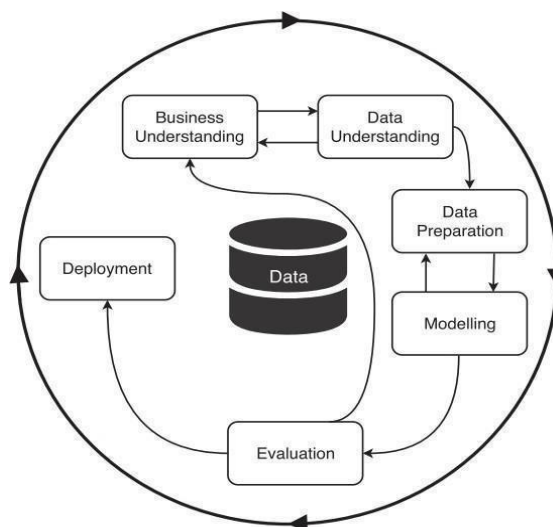


Figura 7. Cross Industry Standard Process for Data Mining (Shearer, 2000, p. 14)

Segundo o pipeline utilizado em ciência de dados, antes de treinar qualquer modelo, torna-se então fundamental a recolha, exploração e transformação dos dados, de modo que haja uma preparação e adequação dos mesmos, aumentando a sua utilidade no processo de ML (Shaw et al., 1993). Não obstante, na saúde, a aplicação de ML requer o pré-processamento, o refinamento dos dados e o treino dos modelos, tornando possível, por último, a apresentação de ilações úteis à obtenção de vantagem competitiva por parte das empresas (Kee Yuan Ngiam, & Ing Wei Khor., 2019).

Apesar de numa primeira impressão a aplicação de IA parecer surgir associada unicamente a aspectos positivos, uma vez que o funcionamento desta tecnologia requer a existência de algoritmos que processam e analisam, em tempo real, uma quantidade enorme de dados de diversas fontes (Mohamed-Ilias et al., 2021), expõem-se algumas críticas sobre os métodos de ML.

A natureza da aplicação de ML é a aprendizagem através de dados históricos e a tomada de decisões sobre novos dados. Isto gera entraves aquando do aparecimento de mudanças imprevisíveis, como é exemplo a recente Pandemia da COVID-19, onde se verificou que as técnicas de ML não tiveram, num momento inicial, o desempenho necessário.

Outra crítica a apontar consiste no facto de a validade e imparcialidade daquelas técnicas ser frequentemente questionada. Em alguns casos, os algoritmos de ML replicam problemas existentes nos dados utilizados para treiná-los (D. D. Cremer., 2020).

2.5 Desafios e benefícios da IA na GCA

À data, as justificações mais plausíveis e que apesar de simples, realçam a necessidade da aplicação dos novos métodos de análise de dados, consistem no facto, como mencionado acima, de os métodos tradicionais não serem capazes de lidar com grandes volumes de dados que não estejam estruturados. Apesar dos obstáculos com que estas se possam deparar, o investimento feito pelas empresas na utilização de IA na saúde não deve, de acordo com a opinião defendida por Whicher, D., & Rapp, T. (2022), deixar de ser considerado, por diversas razões, a sublinhar:

- Um número crescente de países organiza os seus serviços em torno do acesso e da análise aos dados, o que irá aumentar o número e a qualidade das avaliações económicas feitas em saúde associadas às ferramentas de IA;
- A regulamentação da utilização dos dados tem tornado a sua utilização mais segura, quer por parte dos profissionais de saúde, quer pelos utentes que contribuem de forma positiva para a partilha de informações e divulgação das vantagens que lhe surgem associadas.

2.6 Modelos de previsão

A cadeia de abastecimento é definida por Lambert, D. M., Stock, J. R., & Ellram, L. M. (1998), como a integração dos processos de transformação da matéria-prima ao consumidor. A GCA torna-se numa área bastante adequada à aplicação de ML. Algumas vantagens da sua aplicação são, de acordo com Jaipuria, S., Mahapatra, S. S. (2014), a previsão da procura, algo passível de ser realizado através, por exemplo, da análise de dados históricos relativos às entradas de consumíveis clínicos nas Unidades, aos níveis MRP e aos índices de rutura.

Além disso, os modelos preditivos também podem ser usados para prever padrões de utilização de serviços de saúde, permitindo que hospitais e clínicas se preparem, atempadamente, para atender à procura, evitando sobrecargas no sistema, com vista ao aumento da qualidade do atendimento aos pacientes e à melhoria da eficiência operacional (Verhoef et al., 2021; Vial, 2019).

3 Metodologia de investigação

A fim de otimizar a eficiência operacional e de antecipar necessidades futuras da LH, o projeto a desenvolver irá incidir sobre uma metodologia exploratória (Hugo Terras, 2017).

3.1 Técnicas e instrumentos de recolha de dados

As técnicas e os instrumentos utilizados para a recolha de dados são caracterizados tendo em conta a fonte dos dados, fornecidos pelo diretor operacional do CL CUF, em formato Excel.

3.2 Metodologia da Análise dos Dados

A metodologia da análise dos dados adotada será abrangente, incorporando abordagens qualitativas e quantitativas. A realização de três reuniões com os profissionais responsáveis pela gestão de encomendas, gestão de compras e com a central de negociação, a realização de entrevistas informais com a diretora do departamento de Aprovisionamento e Logística Hospitalar da CUF Tejo e com o diretor operacional do centro logístico CL CUF e a observação da metodologia de trabalho daquele departamento, justificam a realização de uma análise qualitativa.

Por sua vez, a análise quantitativa irá envolver a aplicação de inteligência artificial para analisar os dados referentes aos níveis de gestão MRP (gestão de stock por níveis máximos e mínimos), a informações relativas às entradas de materiais no armazém afeto ao Bloco Central Indiferenciado e a dados referentes às ruturas, ambos fornecidos em Excel e exportados para análise com auxílio à linguagem de programação Python.

Por surgir associada à facilidade de utilização, Python terá sido a linguagem de programação escolhida para desenvolver o projeto. A acrescentar, Pandas consistiu na biblioteca de análise de dados, designada para receber dados conforme expõe (Crameri, 2018), fornecidos como um arquivo CSV, e que permitiu a criação de um dataframe. A biblioteca Pandas terá sido utilizada juntamente com o Jupyter Notebook para obter outputs e ter documentação em markdown. Numpy, outra biblioteca utilizada no projeto, terá sido fundamental para a manipulação dos dados. Subjacente à visualização de tabelas e gráficos foram ainda utilizadas as bibliotecas Matplotlib e Seaborn.

O Jupyter Notebook é uma aplicação ou ferramenta web, de código, que permite ao utilizador criar, limpar e transformar dados, fazer modelagem estatística, visualizar dados graficamente e aplicar ML. Após a recolha de dados, a investigação prosseguirá com a manipulação dos mesmos, onde serão preparados, ou seja, limpos, através da remoção de valores nulos, valores duplicados e renomeação dos nomes das variáveis a ter em consideração.

A partir da observação dos dados fornecidos em Excel pelo diretor operacional do CL CUF, será possível verificar a possibilidade de aplicação daqueles dados em ML, bem como a extração de padrões ou tendências nos dados fornecidos (Mehryar Mohri et al, 2018). Este processo será definido como um processo que utiliza a experiência para mitigar o aparecimento de situações de ineficiência, a redundância de operações e as operações sem valor acrescentado, contribuindo ainda, segundo Carvalho, J. C.; Ramos, T. (2013), para melhorar o desempenho e, eventualmente, fazer previsões concretas.

Conforme demonstra a figura 8, o ML, definido como um subdomínio da IA por Ray & Chaudhuri (2021) é definido como o processo através do qual programas de computador processam uma enorme quantidade de dados (*big data*) provenientes de diversas fontes, encontrando padrões que são analisados (Scott, 2021, p. 1388) e que surgem da interseção entre a estatística computacional - relação entre os dados - e a ciência da computação - implementação dos algoritmos na procura de soluções para determinados problemas (Deo, 2015).

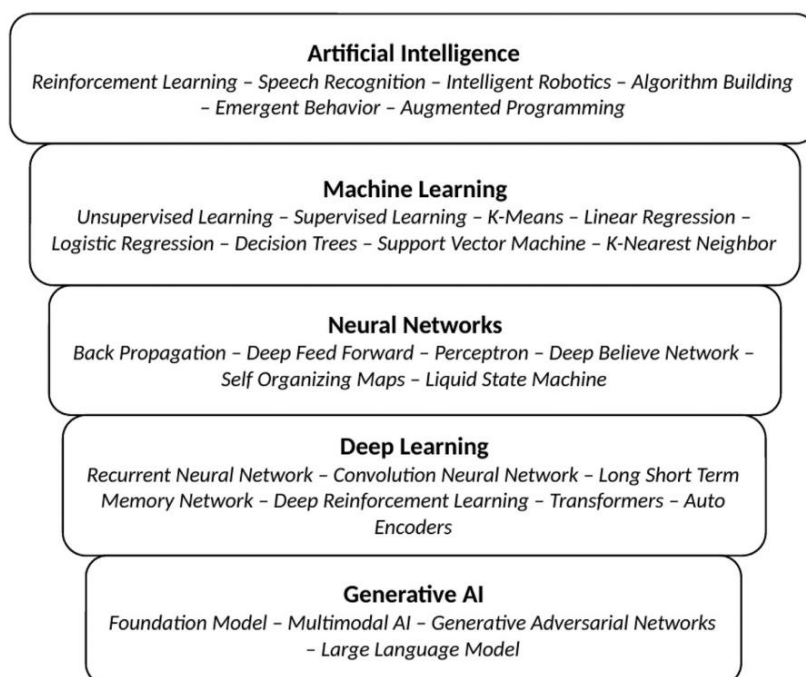


Figura 8. Hierarquia da Inteligência Artificial

4 Logística na CUF Tejo

Conforme exposto na metodologia a adotar, no plano de análise de dados será feita uma análise quantitativa aos dados recolhidos, compreendida através da realização de uma análise prévia, qualitativa e obtida em ambiente hospitalar, da caracterização que é feita aos materiais.

A gestão daqueles consumíveis clínicos, conforme consta na figura 9 e de acordo com a informação exposta por profissionais em contexto hospitalar e através da realização de videoconferências, cabe à LH quando os mesmos se caracterizam por ter elevado custo e pouco consumo, podendo ser geridos como consignações residentes quando adquiridos ao fornecedor e entrando nas contas hospitalares à data do seu consumo. Em contrapartida, os materiais podem também ser geridos como consignações não residentes quando devolvidos ao fornecedor aquando da sua não utilização.

A responsabilidade pelos materiais concerne ao CL CUF quando os mesmos surgem associados a um custo reduzido e constituem stock próprio da Unidade, ou seja, caso se tratem de materiais cujo consumo é elevado e transversal às demais unidades abastecidas pelo CL CUF.

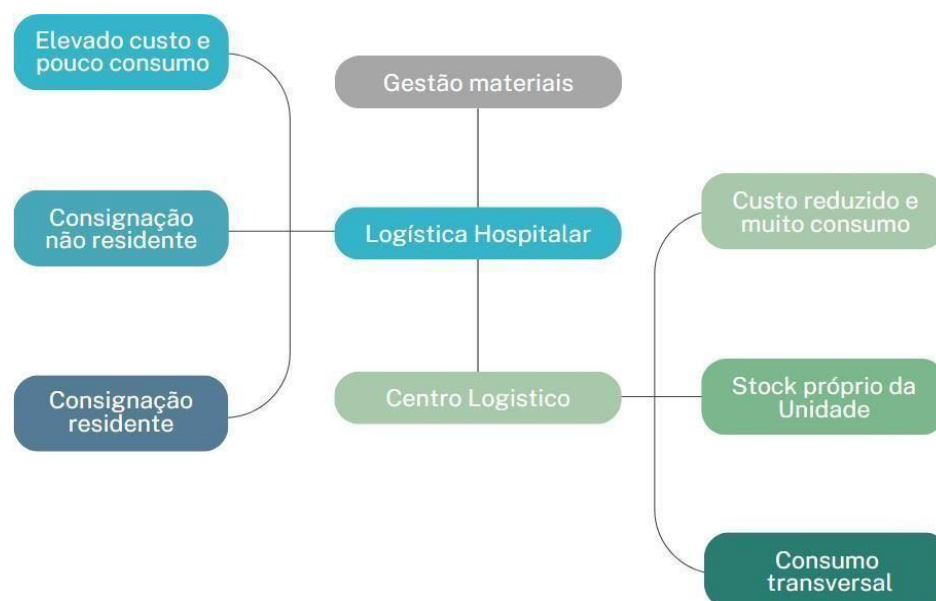


Figura 9. Gestão dos consumíveis clínicos

A logística exposta acima, subjacente à gestão dos consumíveis clínicos, torna possível a existência de um armazém que dá resposta às necessidades diárias do Bloco Operatório.

O layout do BO da CUF Tejo decorre na horizontal e é composto por oito salas cirúrgicas, a nomear, duas híbridas, uma oftalmologia e angiografia e uma de urologia e ginecologia, uma de

neurocirurgia, uma de cirurgia geral, o bloco central indiferenciado, uma de otorrino, uma de oftalmologia, uma de cirurgia robótica.

O armazém do BO da Unidade CUF Tejo, que serve de auxílio às salas de operações mencionadas, organiza-se em 2 pisos, implementados unicamente para a gestão daqueles serviços e abastecidos por fornecedores externos e pelo fornecedor interno CLCUF.

5 Dados

No plano de análise de dados serão consideradas três tabelas exportadas do Excel para um *Jupyter Notebook*, documento onde será utilizada a linguagem de programação *Python* para o sucesso da análise a fazer aos dados. O trabalho exploratório desenvolvido subdivide-se em fases, de modo a reduzir a complexidade que lhe surge associada.

5.1 Recolha

O trabalho desenvolvido incide sobre um estudo experimental, onde são recolhidos dados cuja análise inicial passa pela sua descrição e pela apresentação da forma como os mesmos foram recolhidos.

Os dados fornecidos em Excel, após reunião, pelo diretor operacional do CL CUF, possibilitaram a criação de três tabelas, referentes às entradas dos consumíveis clínicos no serviço alvo de análise, o BCI da Unidade Hospitalar CUF Tejo; aos níveis MRP, através dos quais é feita a gestão dos valores máximos e mínimos associados às encomendas; e aos índices de rutura. A recolha destes dados justifica - se através das mais valias que a realização de uma análise de dados tem na veracidade e coerência das conclusões finais apresentadas.

Conforme exposto no capítulo referente à revisão de literatura, considerando como ideia inicial a aplicação de algoritmos de ML verificou-se, após a obtenção dos dados, que estes não seriam suficientes à aplicação de modelos de previsão. Isto decorre do facto de não ser possível, com os mesmos, extrair informações importantes nomeadamente na escala temporal que relaciona consumos com encomendas.

5.2 Compreensão e preparação

Conforme exposto acima, aquando da verificação da possibilidade de utilização de modelos de classificação, regressão e a posterior avaliação com métricas como *accuracy*, *precision*, *recall*, *confusion matrix* e *f1-score* *mean absolute error* (MAE), *mean squared error* (MSE), *root mean squared error* (RMSE) e *coefficient of determination* (R^2), percebeu-se serem insuficientes as informações passíveis de ser extraídas dos dados. Posto isto, não terá sido possível avançar nesse sentido e procedeu-se apenas com a análise a expor.

A observação das tabelas abaixo permite verificar que existem variáveis cuja análise é de difícil interpretação. A remoção de colunas surge como solução para que seja feita uma análise mais sucinta e específica das variáveis consideradas relevantes. A remoção de valores nulos presentes em colunas e de variáveis cuja análise era de difícil interpretação, promoveu ainda a eliminação de

materiais e evitou o aparecimento excessivo de erros como *outliers* e *missing values* na posterior análise de dados.

A variável referente à data terá sido separada em três novas variáveis, de modo a aumentar a minuciosidade associada à análise temporal. Por último, de forma a tornar simples a leitura das variáveis presentes nos dois quadros à esquerda, estas foram renomeadas, ficando, como visível na figura 9, disponíveis para a análise dos dados, as variáveis apresentadas nos três quadros à direita.

Entradas	Níveis MRP	Entradas	Níveis MRP
Material	Modif. em	Código Sap	Código Sap
Texto	Material	Nome material	Área MRP
Centro	Área MRP	Unidade Hospitalar	Gestão níveis MRP
Departamento	Centro	Serviço BCI	Nível mínimo
Doc.	GMRP	Nº de encomenda	Nível máximo
Itm.	Pt. reabast.	Pedido de compra	Variabilidade
Data lançamento	Est. máximo	Quantidade entrada no BCI	
Quantidade	SuE	Ano de entrada	Índice de ruturas
UMR	DpSuE	Mês de entrada	CUF geral
Data Doc.	Hora Mod.	Dia de entrada	Centro Logístico
Entrado em		Hora de entrada	CUF Tejo
Hora		Qt. UMP	Serviço BCI
Qtd.UMP		Qt. UPP,	
Qtd.UPP		Quantidade	
Quantidade		UMB	
UMB		UMP	
UMP		UPP	
Cen.1		Fornecedor hospitalar	
Fornecedor			

Figura 10. Descrição das variáveis referentes aos dados a analisar

Após a limpeza e transformação dos dados ficaram disponíveis para análise:

- 12266 linhas e 19 colunas afetas aos dados das entradas dos materiais no BCI e na Unidade, provenientes do CL;
- 461 linhas e 4 colunas afetas aos dados referentes aos índices de rutura;
- 546 linhas e 5 colunas nas quais se apresentam os dados referentes às variáveis que permitem a análise dos níveis MRP.

5.3 Análise

Após manipulados, os dados passaram por um novo processo de análise para encontrar correlações, padrões e tendências através da análise de gráficos, tabelas e matrizes, com o intuito de perceber que informações poderiam ser extraídas dos mesmos.

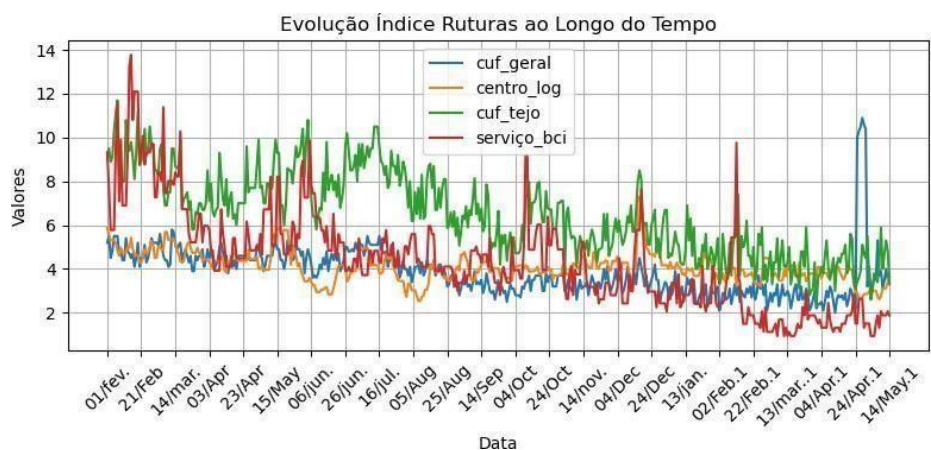
Unnamed: 0	cuf_geral	centro_log	cuf_tejo	serviço_bci
01/fev.	5.2	5.86	9.0	9.33
02/Feb	5.5	5.57	9.5	7.65
03/Feb	4.5	5.30	8.9	5.78
04/Feb	4.8	5.30	9.1	5.78
05/Feb	5.5	5.30	10.3	5.78

Figura 11. Índices de rutura

Após a visualização dos dias e dos valores percentuais do índice de ruturas associado a cada variável, foi utilizado o método *plot()* de *Pandas* para apresentação de um gráfico simples, onde o eixo das abscissas corresponde ao período total analisado e o eixo das ordenadas ao índice de ruturas. A observação dos gráficos abaixo, para os quais podemos considerar um valor de *benchmarking* de seis unidades percentuais, sendo que qualquer dia, associado a cada uma das variáveis cujo índice de ruturas seja superior aquele valor, demonstra resultados negativos e qualquer dia cujo índice de ruturas seja inferior a 6 demonstra resultados positivos. O primeiro gráfico permite aferir que o serviço BCI e o Hospital CUF Tejo são representativos da maior instabilidade na gestão das ruturas, tendo sido excedido em mais do dobro, em fevereiro de 2023 pelo serviço BCI, o valor médio associado ao índice de ruturas. Abaixo, encontram-se descritas algumas ilações aferidas da observação dos quatro gráficos seguintes ao primeiro:

- Apesar de apresentar uma tendência decrescente, algo positivo, até 24 de abril de 2024, a CUF a nível nacional assistiu a um aumento exponencial do seu índice de ruturas nos últimos dias daquele mês, o que sugere uma revisão nos níveis MRP.
- O Centro Logístico tem vindo a diminuir, progressivamente, os seus índices de ruturas, nunca ultrapassando o valor 6 e apresentando valores bastante atípicos no final de dezembro de 2023.
- A gestão logística da CUF Tejo tem apresentado resultados bastante positivos nos últimos 8 meses alvo de análise, em contrapartida aos 7 meses iniciais.
- A dificuldade de controlo sobre fatores externos que influenciam a gestão logística do BCI justificam a irregularidade observada no quinto e último gráfico. Apesar de surgir sobre uma tendência decrescente, algo positivo, aquela irregularidade e a presença de bastantes *outliers*, refletem a importância da apresentação de propostas de melhoria a este serviço indispensável ao sucesso organizacional da Unidade.

O gráfico da evolução do índice de ruturas, sem a apresentação de *outliers* e que decorre entre 1 de fevereiro de 2023 e 14 de maio de 2024 apresenta-se como parte da figura 12, e os gráficos com *outliers* e referentes à evolução do índice de rupturas associado a cada uma das variáveis, apresentam-se nos gráficos abaixo daquele.



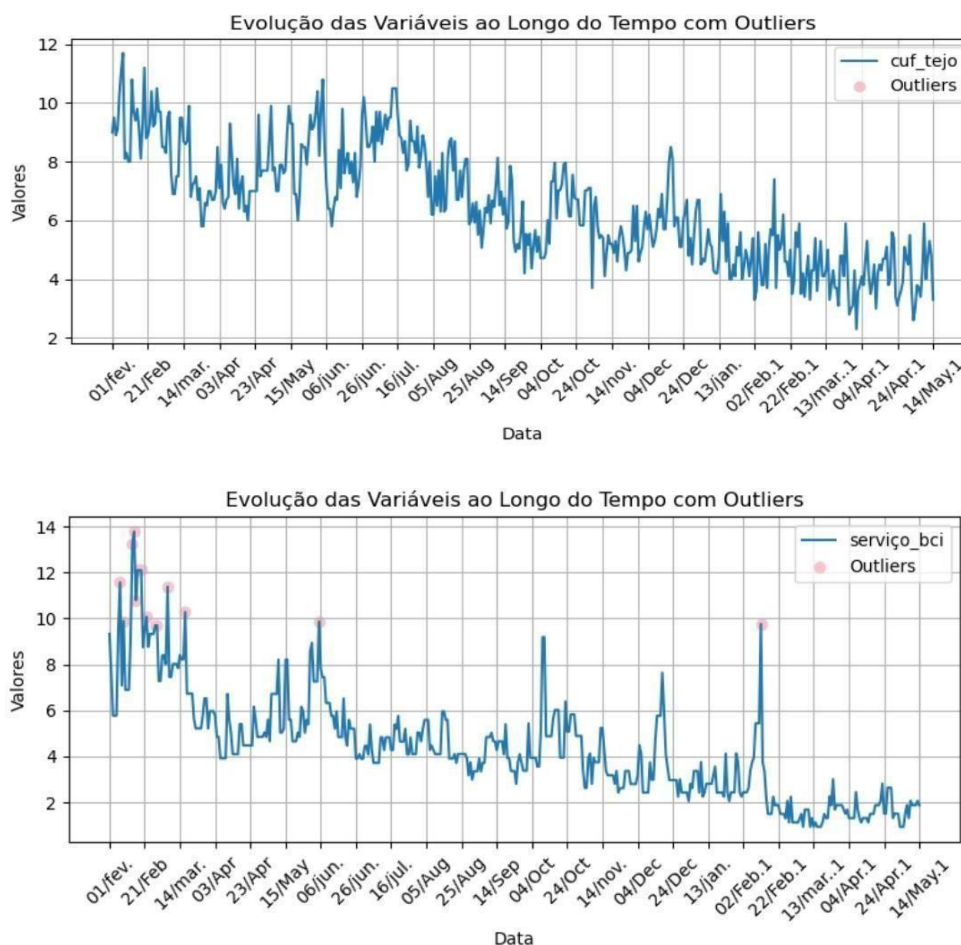


Figura 12. Evolução dos índices de rutura

Subjacente à análise de dados e de acordo com Dancey, C. P., & Reidy, J. (2017), a interpretação do coeficiente de correlação de Pearson é essencial, bem como a compreensão de que correlação não implica causalidade entre as variáveis. Segundo os autores, uma correlação, ou seja, a forma como uma variável se comporta em relação a outra, que seja próxima de 1 indica uma forte correlação positiva e um valor próximo de -1 indica uma forte correlação negativa. Percebe-se que um valor próximo de 0 sugere uma correlação fraca ou inexistente. A elaboração de um *heatmap* permitiu perceber se as quatro variáveis apresentadas são influenciáveis entre si, ou seja, se os serviços prestados em cada uma afetam os serviços prestados noutra. Tendo em conta os resultados obtidos, percebemos que existem duas variáveis cuja correlação é notável e outras duas que funcionam de forma mais autónoma entre si.

A Correlação de Pearson entre a CUF Tejo e a CUF a nível nacional, de 0.56, e entre o BCI e a CUF Tejo, de 0.73, sugerem que o aumento dos índices de rutura alvo de gestão na CUF Tejo e no BCI, promovem um aumento daqueles valores na CUF Geral e na CUF Tejo, respetivamente. A

Correlação de Pearson entre o CL e a CUF Tejo, de 0.36, e entre o CL e a CUF a nível nacional, de 0.13, sugerem que o aumento dos valores alvo de gestão no CL não apresenta uma ligação tão impactante quanto as acima descritas, apesar de condicionar, em cadeia, ambos os índices de rutura do BCI, da CUF Tejo e da CUF Geral.

Conclui-se que uma rutura no serviço BCI se reflete bastante no índice de ruturas do Hospital CUF Tejo, evidenciando a importância de uma correta gestão do seu stock. Por estar afeto somente ao abastecimento daquela Unidade Hospitalar, o índice de ruturas do CL apresenta uma correlação próxima de zero quando estudada a sua influência sobre o índice de ruturas da CUF a nível nacional.

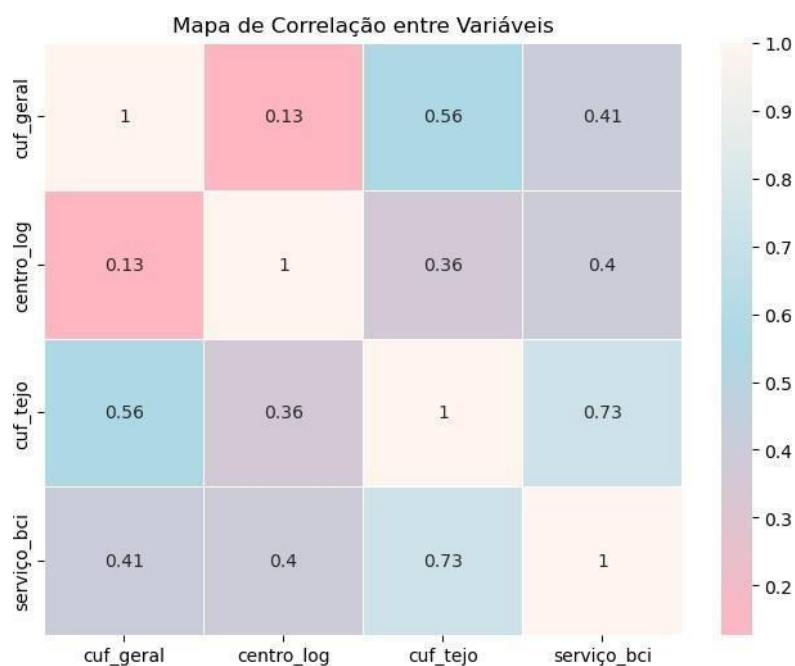


Figura 13. Correlação rutura entre variáveis

A análise de dados através de *scatterplots* caracteriza-se pela facilidade que surge associada à sua visualização (Tory et al., 2007). De modo a visualizar graficamente as correlações apresentadas no *heatmap* acima, os materiais surgem, na figura 14, sob forma de pontos.

O primeiro gráfico de dispersão relaciona as variáveis ‘cuf_tejo’ e ‘cuf_geral’ através de pontos, cada um representativo, como mencionado acima, de um material. Observa-se que o aumento do índice de ruturas afeto à CUF Tejo, é sinónimo de um maior índice de ruturas afeto à CUF Geral. O gráfico apresentado à direita do primeiro relaciona o nível de ruturas do CL com o valor associado à CUF a nível nacional. Nota-se que, tendo em consideração a presença de *outliers*, o aumento do nível de ruturas no CL, não representa de forma notável o aumento daquele valor na

CUF. O terceiro gráfico de dispersão relaciona o nível de ruturas do CL com o valor associado à Unidade CUF Tejo. Não se verifica uma tendência clara de aumento ou diminuição da correlação. Verifica-se ainda um aglomerado de pontos entre os valores percentuais 4 e 5.

De seguida relaciona-se o nível de ruturas do BCI com o valor associado à CUF Tejo. Apesar da presença de outliers, a representação dos pontos sugere uma correlação positiva, na qual o aumento do nível de ruturas no BCI representa o aumento daquele valor na CUF Tejo. O último gráfico relaciona o nível de ruturas do CL com o nível de ruturas do BCI. Por ser notável a presença de bastantes outliers que dificultam a percepção da influência que as ruturas no CL provocam no BCI, sugere-se a remoção dos mesmos.

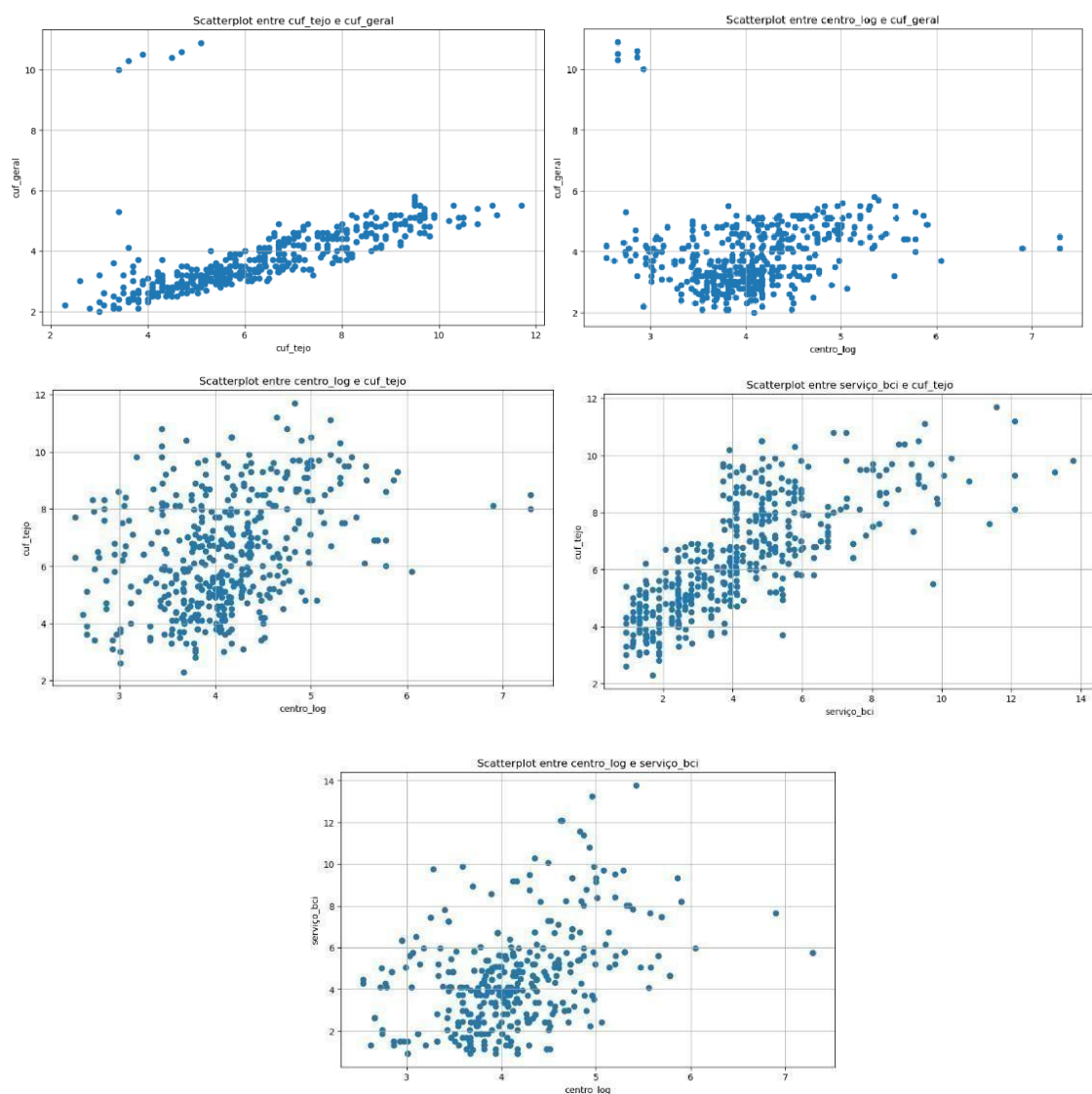


Figura 14. Scatterplots rutura entre variáveis

Na figura abaixo encontram-se 4 boxplots, cada um representativo da distribuição estatística de cada uma das variáveis em análise. Percebe-se que o nível de ruturas da CUF Tejo possui como mediana um valor próximo de 7, superior a qualquer outra, apesar da presença de *outliers* nas restantes, cujas medianas rondam as 4 unidades percentuais. A acrescentar, a CUF Tejo apresenta também uma maior dispersão dos dados. O BCI, apesar de apresentar mais valores atípicos, não considerados nos bigodes do gráfico, apresenta o valor mínimo mais pequeno, algo que pode ser representativo da sua importância e do controlo acrescido que é tido sobre o possível aumento do nível associado à rutura de consumíveis afetos aquele serviço do BO.

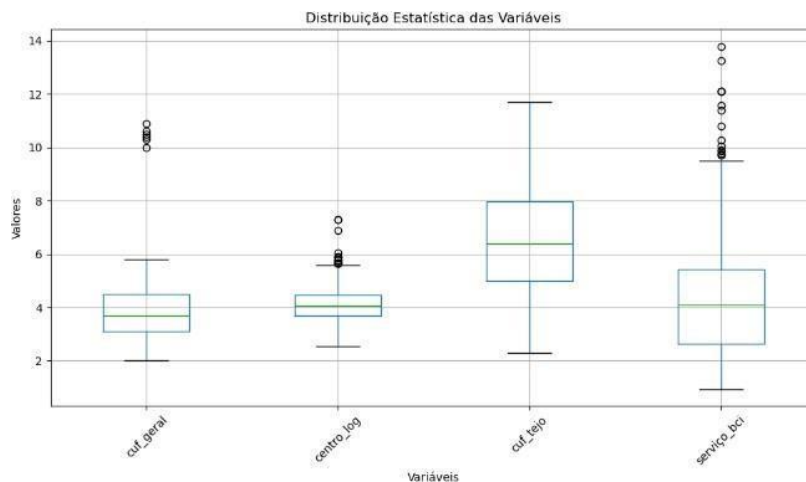


Figura 15. Distribuição estatística das variáveis

O gráfico 16 permite-nos aferir que quanto menor o nível de reabastecimento de stock, maior o número de vezes que se verifica qualquer quantidade a que é dada entrada no BCI, sendo preciso reajustar quantidades de stock com mais regularidade. Para as quantidades mais utilizadas no BCI, são estipulados níveis MRP inferiores de modo que a reposição de stocks ocorra com maior regularidade. A quantidade “entrada” deve ser sempre superior ao nível mínimo MRP. A maioria dos dados está concentrada nos valores mais baixos de Nível Mínimo MRP. Isso sugere que a maior parte das entradas de material no BCI corresponde a níveis mais baixos de MRP.

Existe uma alta densidade de pontos entre 0 e 500 para o Nível Mínimo MRP. A presença de outliers é visível, especialmente para valores mais altos de Nível Mínimo MRP, representativos de materiais para os quais a quantidade de entrada no BCI é alta ou baixa, apesar de altos níveis mínimos MRP. Apesar das conclusões aferidas através da observação do gráfico, a correlação entre aquelas variáveis aproxima-se de 0, o que sugere que o nível mínimo MRP não é o único fator que afeta a quantidade de entradas no BCI.

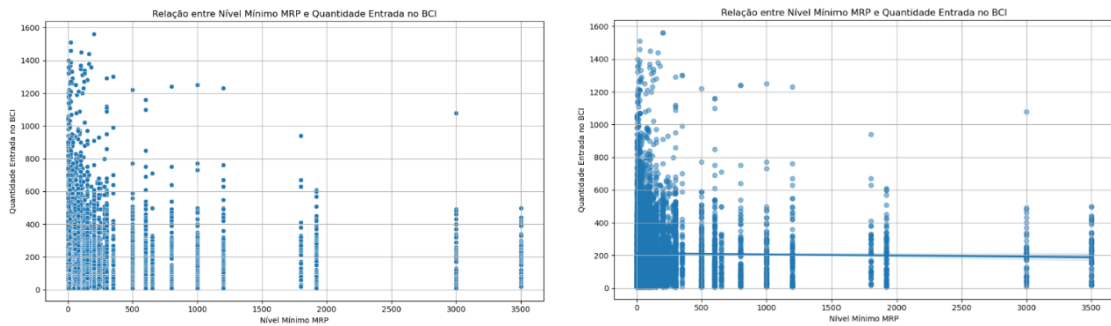


Figura 16. Relação Nível Mínimo MRP e Quantidade Entrada no BCI

O gráfico 17, representativo da distribuição da quantidade de consumíveis clínicos entrada no BCI ao longo do tempo, permite aferir que entre setembro e outubro de 2023, período no qual se recorreu a uma revisão dos parâmetros logísticos do BO, ocorreu um aumento da quantidade de entradas no BCI. Seis meses depois assistiu-se a uma tendência inversa, com a redução do número de materiais entrados, algo que pode refletir-se como uma consequência positiva da revisão dos seus parâmetros, possivelmente o correto ajuste dos níveis MRP e o controlo do excesso de stock nos AA. Os valores extremos observados justificam a sua razão de ser por estarem associados aos picos de quantidades entradas, de diversos materiais, no intervalo de um dia.

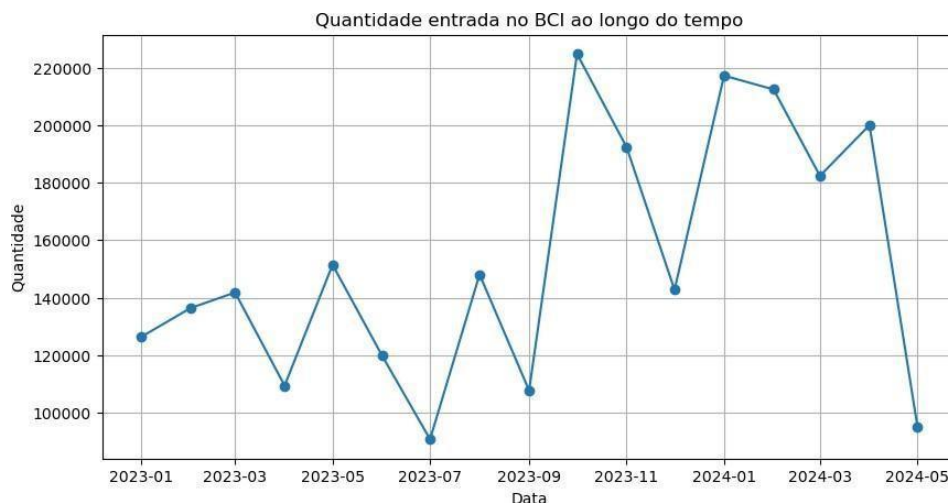


Figura 17. Quantidade entrada no armazém do BCI

A análise ABC, técnica utilizada na gestão de stocks com o intuito de categorizar materiais em categorias pré-definidas, baseia-se no Princípio de Pareto para determinar que materiais devem ter prioridade aquando da sua gestão (Douissa, M. R., & Jabeur, K., 2016). Os materiais com maior relevância, que constituem 5% a 10% do stock pertencem à categoria A. Os itens que constituem

20% a 30% são classificados na categoria B e itens considerados como menos importantes (50% a 70%) constituem a categoria C.

Tendo por base a variável referente ao número de vezes que foi dada entrada a consumíveis clínicos afetos ao BCI no intervalo de tempo estudado, e com o objetivo de, a partir da utilização desta técnica de gestão de stocks, ser possível determinar qual a política de controlo de inventário a usar para cada classe, procedeu-se à realização do gráfico abaixo, representativo da distribuição dos consumíveis clínicos pelas classes que lhes correspondem. A análise ABC realizada permite inferir que a literatura existente é corretamente aplicada, distribuindo-se, de forma decrescente, as quantidades a que é dada entrada no armazém do BCI.

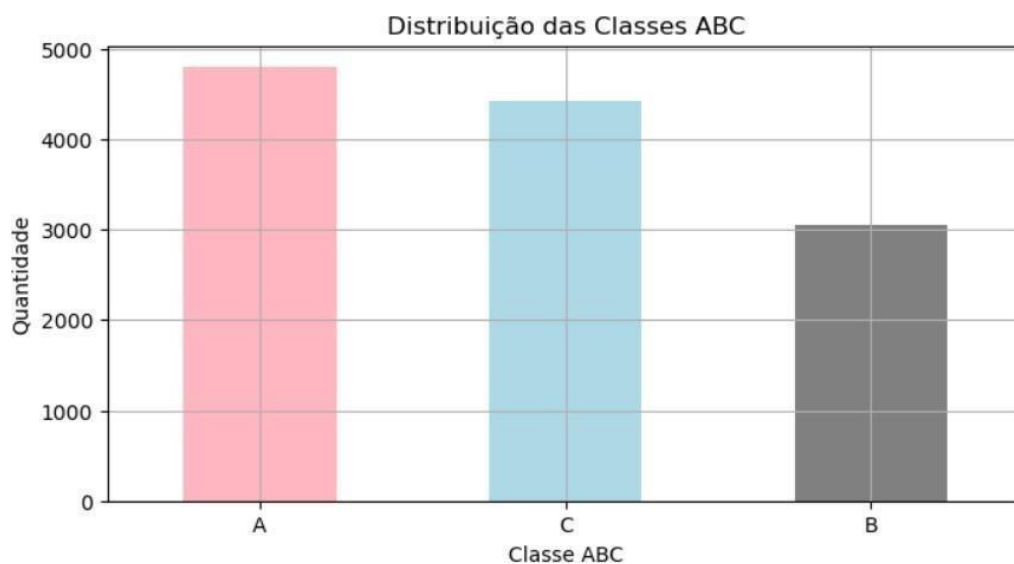


Figura 18. Distribuição dos materiais por classe ABC

Conforme apresentado abaixo, de modo a clarificar a funcionalidade daquela aplicação, os dados disponíveis possibilitaram ainda a identificação do top 5 de materiais associados a cada uma das classes ABC.

classe_abc	cod_sap	nome_material
A	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA
A	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA
A	110000746	SISTEMA INFUSÃO FLUIDOS BIPERFURADOR
A	110000369	MASCARA OXIGENIO ADULTO
A	110000321	OCULO NASAL ADULTO
B	110000118	PENSO NÃO TECIDO 9 X 15 CM
B	110026349	OBTURADOR BIDIRECIONAL
B	110006885	MASCARA LARINGEA I-GEL Nº3.0
B	110002997	BATA CIRURGICA REFORÇADA TAM XL
B	110003616	SUORTE BRAÇO ADULTO
C	110000971	LIGADURA ELASTICA FORTE COMPRESSÃO 10 CM
C	110000012	PENSO RAPIDO ADULTO
C	110000499	CATETER ENDOVE PERIFERICO PROTEC14GX45MM
C	110004555	LIGADURA ELASTICA AUTO-ADERENTE 15 CM
C	110000680	REFERENCIA VASCULAR VERMELHA

Figura 19. Top 5 materiais de cada classe

6 Resultados

Uma vez terminada a análise reportada acima, terão sido averiguadas quais as transformações a ser feitas para atingir os resultados através, por exemplo, da criação de uma nova variável.

De acordo com os dados disponíveis e tendo em conta os objetivos da investigação, sugere-se a criação de uma variável binária, o máximo variável.

Ao sugerir a criação de uma nova variável, útil à gestão dos stocks do BCI, percebe-se necessária a criação de outras variáveis, como o número de materiais encomendados entre janeiro de 2023, ano no qual foram encomendados 8209 materiais e abril de 2024, ano no qual foram encomendados 4057 materiais.

Após a criação da variável acima, sugere-se a criação de uma outra, representativa do número total de encomendas mensais por código SAP, ou seja, por material. Verificou-se que a média do número total de encomendas não inclui materiais para os quais não tivessem sido realizadas quaisquer encomendas. Sendo que a média obtida, de cerca de 2.43, ou seja, aproximadamente 2 encomendas, não corresponde à realidade, por excluir, como indicado, meses onde determinado material surge sem a colocação de qualquer encomenda, terão sido adicionados, ao cálculo da média, os materiais sob aquela condição. Ao incluí-los, a média mensal de encomendas realizadas por material diminuiu, passando a 1.37, ou seja, cerca de uma encomenda associada, mensalmente, a cada material. De sublinhar, que uma encomenda de material, pode possuir diversas unidades daquele consumível clínico que, por sua vez, podem ser geridos à caixa ou à embalagem.

Tendo em conta os 525 materiais alvo de encomenda, antes de prosseguir, foram analisados alguns dados estatísticos:

- O material mais vezes presente numa encomenda, sempre nas quantidades necessárias à satisfação do seu nível MRP, foi solicitado ao fornecedor CL catorze vezes, no período de um mês.
- Determinado material foi encomendado, mais uma vez, nas quantidades necessárias à satisfação do seu nível MRP, em média, cerca de 1.37 vezes.
- O desvio padrão, ou seja, a dispersão dos valores em relação à média, foi de 1.7.

A soma do valor do desvio padrão com a média de encomendas, por material, realizadas mensalmente, igualou um valor de referência, nomeado limite e que servirá de base à gestão do nível máximo MRP. Assim, com a finalidade de definir os materiais a analisar, terão sido filtrados apenas os materiais cuja quantidade mensal de encomendas supera o valor obtido naquele limite, ou seja, o valor de referência.

De modo a ajustar as quantidades encomendadas e, eventualmente, o nível máximo associado ao nível MRP de cada material, sugere-se a realização de uma análise experimental com mais dados, na qual qualquer material cuja quantidade encomendada mensalmente seja superior ao valor de referência, deva sofrer um aumento de duas vezes o seu nível máximo de reposição, passando a ser caracterizado por ter um máximo variável. Esta nova variável de apoio à gestão irá dar resposta a hipotéticas necessidades no mês seguinte. Caso se verifique o cenário contrário nos meses seguintes, o nível máximo volta a ser adaptado, evitando o aparecimento de excesso de stock no armazém afeto ao BCI.

	cod_sap	year_entrada	month_entrada	total_month_orders
0	110000012	2023	1	2
1	110000012	2023	2	0
2	110000012	2023	3	1
3	110000012	2023	4	0
4	110000012	2023	5	1
5	110000012	2023	6	1
6	110000012	2023	7	0
7	110000012	2023	8	2
8	110000012	2023	9	1
9	110000012	2023	10	1
10	110000012	2023	11	1
11	110000012	2023	12	3
12	110000012	2024	1	2
13	110000012	2024	2	1
14	110000012	2024	3	1
15	110000012	2024	4	2
16	110000012	2024	5	0

Figura 20. N° de presenças mensais de determinado cod_sap numa encomenda

A atuação da CUF traduz o seu empenho contínuo em assegurar a criação de valor em benefício da satisfação dos clientes, colaboradores, acionistas e outras entidades com as quais colabora no exercício da sua atividade. A melhoria contínua do desempenho global de uma organização deve ser um objetivo permanente dessa organização. De modo a aperfeiçoar, tornando mais eficiente o ajuste dos níveis MRP, sugere-se a criação de uma nova variável, passando a gestão dos stocks do BCI a ser feita tendo em conta a existência de um valor limite.

Resumindo, dada a equação:

$$\text{Limite} = \text{Desvio Padrão} + \text{Média de Encomendas por Material por Mês} \quad (1)$$

Definindo como condição, se a quantidade mensal de encomendas realizadas para repor determinado consumível clínico for maior que aquele limite, então aplica-se a variável binária máximo variável, onde 1 indica que aquele limite foi excedido, sugerindo-se, nestas situações, o aumento do nível máximo para o dobro do seu valor atual e 0 indica que o limite não é excedido. Nesta última situação, o nível máximo não sofre qualquer ajuste, permanecendo consoante o nível máximo definido no mês anterior para o mês presente. O máximo variável adicionado, na última coluna da tabela abaixo, e para a qual 1 representa os cod_sap, associados aos materiais, para os quais a quantidade encomendada excede o limite estipulado, permite ainda aferir que existem 1058 materiais cujo nível máximo deve ser ajustado, no mês vindouro pela equipa responsável pela gestão do armazém do BCI. Por outro lado, a diferença entre a quantidade de encomendas realizadas e aquele valor, indica-nos que 11208 materiais se encontram abaixo do limite estipulado para o número de encomendas a realizar mensalmente consoante o nível máximo definido no mês anterior para o mês presente.

	cod_sap	year_entrada	month_entrada	total_month_orders	max_variavel
37	110000018	2024	2	6	1
60	110000020	2024	1	7	1
61	110000020	2024	2	4	1
121	110000025	2023	2	4	1
122	110000025	2023	3	5	1
...	...	...	...	...	...
12491	110088496	2023	12	5	1
12492	110088496	2024	1	5	1
12493	110088496	2024	2	7	1
12494	110088496	2024	3	6	1
12495	110088496	2024	4	5	1

1058 rows × 5 columns

Figura 21. Variável binária

7 Discussão de resultados

O presente trabalho teve como objetivo primordial a compreensão da importância que tem a análise dos padrões do fluxo do grupo de consumíveis clínicos afetos ao bloco central indiferenciado, cujo processo logístico implica a sua passagem pelo CL CUF. Com vista ao cumprimento do objetivo, terão ainda sido apresentados objetivos secundários. Para responder ou cumprir o primeiro objetivo secundário, realizou-se uma análise detalhada aos valores associados aos índices de rutura daqueles materiais, no período de um ano e quatro meses, aos níveis MRP e às variáveis associadas às entradas de material na Unidade hospitalar.

O cumprimento do segundo e do terceiro objetivos também terá sido possível. A análise aos dados recolhidos permitiu a seleção de variáveis cuja limpeza e transformação tornou possível a realização de gráficos e a identificação de ineficiências na gestão de stocks. Aqueles gráficos tornaram mais fácil a avaliação do consumo e do abastecimento do grupo de materiais alvo de estudo, a realização de previsões relativas ao futuro, e obter conclusões, como é exemplo a possibilidade de aumentar o nível máximo para o dobro do seu valor atual, sugerindo consequentemente a diminuição do número de vezes que a reposição do material é realizada na prateleira do Armazém Avançado.

Deste modo, verificou-se o cumprimento do último objetivo secundário, tendo sido encontradas soluções para resolver dificuldades que ocorreram no período de tempo em questão e aquando da operação logística.

8 Conclusão

Os profissionais logísticos demonstram as suas competências em situações de crise, através do modo como são capazes de responder às necessidades (VanVactor, J., 2011). Apesar de ser um processo de correção e melhoria constante, a gestão decorre em situações nas quais se verifique que as ineficiências existem e mitigam os pontos fortes das empresas, até serem geridas ou resolvidas corretamente.

O objetivo proposto aquando da seleção do tema a estudar ao longo do trabalho terá sido cumprido por ser possível compreender que a análise de dados históricos referentes aos padrões de consumo e abastecimento de consumíveis clínicos afetos ao BCI contribui para a gestão do abastecimento do CL à Unidade Hospitalar CUF Tejo. Sugeriu-se a criação de uma nova variável, de modo a aprimorar o fluxo de colocação de encomendas já existente e a dar resposta às necessidades futuras de abastecimento de materiais de consumo clínico, com a apresentação de uma proposta de melhoria para a gestão dos stocks da Unidade.

A adoção de uma metodologia que permitisse compreender a logística existente e propor soluções, facilitou o cumprimento do objetivo definido, de compreender a importância que tem a análise dos padrões de consumo e abastecimento associados ao grupo de consumo clínico afetos ao Bloco Central Indiferenciado e cujo processo logístico implica a sua passagem pelo CL CUF.

A realização da revisão de literatura e as informações partilhadas pelos profissionais do Hospital CUF Tejo permitiram fazer ilações relativas aquela que é a realidade da logística hospitalar desta unidade e às representações gráficas obtidas.

O ficheiro de Excel, composto por três pastas, onde constam as informações necessárias à análise dos dados nelas presentes permitiram a realização do trabalho que teve, numa fase inicial, por base, as 560 linhas referentes aos níveis MRP, as 12267 linhas acerca das informações sobre as entradas no armazém designado para servir o Bloco Central Indiferenciado e as 462 linhas referentes às ruturas verificadas no Armazém Avançado, nelas presentes.

A análise destes dados tornou possível a realização de uma análise temporal da evolução do índice de ruturas, relativo a diversas variáveis, e da distribuição da quantidade de consumíveis clínicos entrada no BCI ; a criação de boxplots, cada um representativo da distribuição estatística de cada uma das variáveis em análise; a verificação da existência de correlações entre as mesmas; a realização da análise ABC tendo por base as quantidades a que é dada entrada no armazém do BCI; a identificação do top 5 de materiais associados a cada uma das classes ABC; e a identificação de motivos que justifiquem lapsos verificados nos dados.

O método de trabalho adotado permitiu, além do exposto acima, a identificação das regras logísticas existentes à data e a apresentação de novas regras, como a definição de uma condição onde, se a quantidade mensal de encomendas realizadas para repor determinado consumível clínico for maior que a soma do desvio padrão com a quantidade média mensal de encomendas, então aplica-se a variável binária nomeada máximo variável, onde 1 representa os “cod_sap”, associados aos materiais, para os quais a quantidade mensal de encomendas realizadas para cada um deles excede o limite estipulado e 0 indica que o limite não é excedido.

Sendo a consequência da condição apresentada, o aumento do nível máximo para o dobro do seu valor atual, existem 1058 materiais cujo nível máximo deve ser ajustado, no mês vindouro, pela equipa responsável pela gestão do armazém do BCI e 11208 materiais que se encontram abaixo do limite estipulado para o número de encomendas a realizar mensalmente, devendo ser mantido o seu nível máximo.

8.1 Implicações do trabalho para a gestão

Apesar das suas vantagens, na prática, a aplicação de técnicas de Machine Learning na CUF, bem como em qualquer outra Unidade prestadora de cuidados de saúde, requer tempo por parte dos profissionais, para estarem presentes em formações. Essas formações são essenciais para que os profissionais compreendam os fundamentos teóricos subjacentes às técnicas de Machine Learning, bem como os potenciais impactos na prática clínica.

8.2 Limitações

No decorrer da elaboração do trabalho identificaram-se algumas limitações, entre elas, a dificuldade ou até impossibilidade de controlar as ruturas de stock existentes a nível mundial e o aumento do número de ocorrências cirúrgicas.

Outra limitação consistiu no facto de não serem fornecidos dados referentes a variáveis que podiam aperfeiçoar a análise feita e as conclusões, como são exemplo os custos associados aos materiais analisados e as datas do consumo dos mesmos. A disponibilidade e utilização destas datas poderia ter sido sinónimo da possibilidade de definição de uma variável binária e da aplicação de modelos de previsão. Aquela variável target, binária, tornaria possível a realização de um *test/train split* e indicar que perante determinada condição poderia ocorrer rutura de stock no BCI, correspondendo 1 à ocorrência de rutura e 0 à não ocorrência de rutura.

Conforme o nome indica, uma rutura ocorre quando se verifica, a curto prazo, a falta de determinado material. Deste modo, estão reunidas as condições para a rutura de determinado stock

quando a quantidade disponível no AA, após o consumo de determinado material, é inferior ao nível mínimo estipulado. Por sua vez, a criação de uma variável target permitiria detectar a ocorrência de rutura caso o intervalo entre a data do pedido de compra e a data da quantidade consumida fosse muito inconstante. A partir disto seria possível sugerir à Unidade, após a avaliação de métricas adequadas, como a *accuracy* - proporção de previsões corretas em relação ao total de previsões (ao ser treinado, o modelo aprende padrões nos dados e usa esse “conhecimento” para fazer previsões ou inferências sobre novos conjuntos de dados) - e a *precision* - proporção de verdadeiros positivos (TP) sobre a soma de verdadeiros positivos e falsos positivos (FP) - um ajuste fundamentado dos níveis MRP ou a automatização da colocação de notas de encomenda para dar resposta às necessidades.

Apesar de, conforme descrito ao longo do projeto, a aplicação de IA nas Unidades prestadoras de cuidados saúde seja cada vez mais uma realidade alcançável, necessária e promotora da obtenção de vantagem competitiva, para que este impacto se torne efetivo, um relatório publicado na Forbes Portugal sugere que o país garanta a criação de um ambiente favorável à inovação e a diminuição do défice de competências. O relatório permite ainda aferir que o acesso à aplicação das mais recentes tecnologias, poderá resultar num crescimento de 61 mil milhões de euros na economia portuguesa (Dírcia Lopes, 2024).

8.3 Sugestões

As tecnologias de informação são essenciais para promover uma cadeia de valor na saúde mais eficiente, centrada no utente e aperfeiçoada com base em evidências, através da análise de dados.

Associados ao valor na saúde surgem, como elementos fundamentais, os utentes, que esperam resultados que correspondam aos custos, os profissionais, que prezam pela satisfação dos pacientes e pela eficiência dos serviços de saúde, os acionistas que pretendem um investimento sem riscos associados e os fornecedores de consumíveis clínicos, que primam pelo equilíbrio entre os preços de venda e as relações com a Unidade. Posto isto, e de acordo com Porter M. A. (2010), denota-se que na área da saúde, o conceito de valor possui diversas conotações, o que dificulta a sua compreensão e utilização.

Associado aos cuidados de saúde baseados no valor está, na ótica do utente, o sucesso de um tratamento medido não apenas pela melhoria dos resultados clínicos, mas também pela perceção que o doente tem da sua qualidade de vida após o receber. A par da perceção do valor em saúde tida pelo paciente, existe ainda a perspetiva considerada pela Unidade prestadora dos cuidados de saúde. Aqui, este conceito pode ser definido como a diferença entre os benefícios percebidos pelos

clientes no momento da aquisição dos produtos ou serviços e os custos económicos para a empresa que os produz ou entrega. Na ótica da Unidade, de acordo com Peteraf M. A. (2003), valor equivale ao excedente, conceito económico representativo da soma dos excedentes do produtor e do consumidor.

8.3.1 Sugestões para investigação

Na primeira década de implementação do artigo 64.º da Constituição da República Portuguesa, que garantiu, a partir de 1976, o direito à saúde em Portugal a todos os cidadãos, os profissionais ajudaram a resolver os problemas de saúde sem ter em conta os custos, tornando perceptível que esta abordagem se tornaria insustentável a curto e médio prazos (Moniz e Barbosa, 2008). Esta questão tornou-se mais complexa com o aumento da esperança média de vida (Instituto Nacional de Estatística, 2017) e consequente aumento da procura por cuidados de saúde, devido ainda à transição epidemiológica, na qual a incidência de doenças infecciosas (transmissíveis) diminuiu, aumentando a incidência de doenças crónicas (não transmissíveis), (WHO, 2017) associadas a tratamentos prolongados e custos elevados. Posto isto, torna-se compreensível a crescente despreocupação das Unidades com a prevenção da saúde, substituída pela preocupação com os custos afetos à prestação de cuidados aos utentes. Deste modo, é possível afirmar que se tem vindo a sobrevalorizar a ótica de valor tomada pelas Unidades ao valor associado à saúde considerado pelos utentes.

Tendo por base o argumento apresentado acima e em conta que o decorrer da elaboração do projeto foca o acréscimo de valor na saúde através de melhorias na gestão logística da Unidade CUF Tejo, será certamente interessante, em trabalhos futuros, explorar literatura referente ao valor em saúde tido em conta pelos pacientes, de modo a verificar, por exemplo, que alterações ou sugestões de melhoria podem ser propostas às Unidades prestadoras de cuidados de saúde, para que as mesmas trabalhem com vista ao desenvolvimento de um maior equilíbrio, quando se aborda o valor na saúde considerado por ambas as partes.

8.3.2 Sugestões para o Hospital

A acrescentar, acredito ser do interesse de qualquer Unidade prestadora de serviços de saúde, a introdução de profissionais formados em informática e gestão de empresas para que, em conjunto com os profissionais formados em gestão de serviços de saúde, em medicina e em enfermagem, reúnam ideias e promovam o desenvolvimento e a implementação de métodos de gestão mais profundos através, como exposto, da análise de dados históricos, de modo a que se torne cada vez

numa realidade mais aceite, a realização de previsões que possam dar resposta a variadas necessidades cuja premência seja a curto prazo, como é exemplo a reposição de stocks.

9 Referências

- Aguilar-Escobar, V.G., Bourque, S. & Godino-Gallego, N. (2015). Hospital kanban system implementation: Evaluating satisfaction of nursing personnel. *Investigaciones Europeas de Direccion y Economia de la Empresa*, 101–110. <https://doi.org/10.1016/j.iedee.2014.12.001>
- Ashwani Kumar, Venkatesh Mani, Vrandra Jain, Himanshu Gupta, & V.G. Venkatesh. (2023). Managing healthcare supply chain through artificial intelligence (AI): A study of critical success factors. *Elsevier*, 13–14. <https://doi.org/10.1016/j.cie.2022.108815>
- Biggio B., & Roli. F. (2018, October). Wild patterns: Ten years after the rise of adversarial Machine Learning. *Pattern Recognition*, 84, 317-331). <https://doi.org/10.1016/j.patcog.2018.07.023>
- Beaulieu, M., & Bentahar, O. (2021). Digitalization of the healthcare supply chain: A roadmap to generate benefits and effectively support healthcare delivery. *Technological Forecasting and Social Change*, 167. <https://doi.org/10.1016/J.TECHFORE.2021.120717>
- Bowles. (2015). *Machine Learning in Python: Essential Techniques for Predictive Analysis*, John Wiley and Sons, Hoboken, NJ, USA.
- Cátia Mateus, João Carlos Santos, & Salomé Rita. (2023). Ricardo Baptista Leite, CEO da HealthAI: “Não se metam na política sem a independência de uma profissão que vos permita dizer ‘não!’” *Expresso*.
- Chalotra, V. (2013). Inventory Management and Small Firms Growth: An analytical study in supply chain. *Vision*, 17(3), 213–22. <https://doi.org/10.1177/0972262913496726>
- Carvalho, J. C., & Ramos, T. (2013). *Logística na Saúde*. Lisboa: Edições Sílabo.
- Crameri, F. (2018). Geodynamic diagnostics, scientific visualization and StagLab 3: Geoscientific Model Development, 11, 2541-2562.

D. D. Cremer. (2020). What Does It Really Take to Build a Fair AI? Harvard Business Review, Massachusetts, MA, EUA.

Dancey, Christine & Reidy, John. (2017). Statistics Without Maths for Psychology (7th edition).

Deo, R. C. (2015). Machine learning in medicine. *Circulation*, 132(20), 1920–1930.
<https://doi.org/10.1161/CIRCULATIONAHA.115.001593>

Dey. (2016). Machine learning algorithms: a review. *International Journal of Computer Science and Information Technologies*, 7(3), 1174-1179.

Dírcia Lopes. (2024, February). Forbes Portugal: IA Pode Resultar Num Crescimento de 61 Mil Milhões de Euros Na Economia Portuguesa.

Douissa, M. R., & Jabeur, K. (2016). A New Model for Multi-criteria ABC Inventory Classification: Proaftn Method. *Procedia Computer Science*, 96, 550–559.
<https://doi.org/10.1016/J.PROCS.2016.08.233>

DP World. (2024). Economic Impact. Trade in Transition 2024.

Egger, J., & Masood, T. (2020). Augmented reality in support of intelligent manufacturing—a systematic literature review. *Computers & Industrial Engineering*, 140.
<https://doi.org/10.1016/j.cie.2019.106195>

Erich Gamma, Richard Helm, Ralph E. Johnson, and John M. Vlissides. (1993). Design Patterns: Abstraction and Reuse of Object-Oriented Design. In *Proceedings of the 7th European Conference on Object-Oriented Programming (ECOOP '93)*, 406–431.

Fernandes, A. C. (2022). Saúde em Portugal. Editora D'Ideias.

Gardas, B. B. (2022). Organizational hindrances to Healthcare 4.0 adoption: An multicriteria decision analysis framework. *Journal of Multi-Criteria Decision Analysis*, 186–195.
<https://doi.org/10.1002/mcda.1766>

Habebh, H., & Gohel, S. (2021). Machine Learning in Healthcare. *Current Genomics*, 22, 291. <https://doi.org/10.2174/1389202922666210705124359>

Hollnagel, E. (2017). Why is work-as-imagined different from work-as-done? *Resilient Health Care*, 2, 279-294. <https://doi.org/10.1016/j.apergo.2023.104209>

Instituto Nacional de Estatística. www.ine.pt, acedido em 2024

Jaipuria, S., & Mahapatra, S. S. (2014). An improved demand forecasting method to reduce bullwhip effect in supply chains. *Expert Systems with Applications*, 41(5), 2395-2408. <https://doi.org/10.1016/j.eswa.2013.09.038>

Jan A Hazelzet. Can fuzzy logic make things clearer? (2009). *Critical care*, 13(1), 1-2. <https://doi.org/10.1109/ICOEI53556.2022.9776844>

Jorge Nascimento Rodrigues. (2024). Crise no Mar Vermelho agrava-se. Mais de 500 navios desistem da rota pelo Suez. *Expresso*.

Jörg Genner., Ștefan-Rhklégas (2024). Relatório Especial Ambições da UE para a inteligência artificial: Melhor governação, investir mais e com mais orientação: as chaves do futuro European Court of Auditors, 8–12.

Kee Yuan Ngiam, & Ing Wei Khor. (2019). Big data and machine learning algorithms for health-care delivery. <https://pubmed.ncbi.nlm.nih.gov/31044724/>

Kotsiantis, S.B. (2007). Supervised Machine Learning: A Review of Classification Techniques, 31, 249-268.

Kwon, I. W. G., Kim, S. H., & Martin, D. G. (2016). Healthcare supply chain management; strategic areas for quality and financial improvement. *Technological Forecasting and Social Change*, 113, 422-428. <https://doi.org/10.1016/j.cie.2022.108815>

Lambert, D. M., Stock, J. R., & Ellram, L. M. (1998). *Fundamentals of logistics management*. McGrawHill.

Lopes, Americo Jorge Marinho. (2012). Gestão do Bloco Operatório. Universidade do Minho, Escola de Economia e Gestão.

Machry, H., Joseph, A., Wingler, D. (2021). The fit between spatial configuration and idealized flows: mapping flows in surgical facilities as part of case study visits. 14(1), 237–250.
<https://doi.org/10.1177/1937586720928350>

Mahraz Mohamed-Ilias, Benabbou Loubna, & Berrado Abdelaziz. (2021). Is Machine Learning Revolutionizing Supply Chain? <https://doi.org/10.1109/GOL49479.2020.9314713>

Mary Shaw and David Garlan. (1996). Software Architecture: Perspectives on an Emerging Discipline. Prentice-Hall, Inc., USA.

Mehryar Mohri, Afshin Rostamizadeh, Ameet Talwalkar. (2018). Foundations of Machine Learning (Second Edition). MIT Press.

Molina, Arturo. (2023). Transforming the Healthcare Supply Chain with Artificial Intelligence. Advanced Algorithms! | LinkedIn. Retrieved July 1, 2024

Monden, Y. (2012). Toyota Production System: An Integrated Approach to Just-In-Time, 4^a Ed., CRC Press. <https://doi.org/10.1201/b11731>

Moniz, H., e Barbosa, C. (2008). Legislação de Direito da Medicina. (C. Editora, Ed.).

Peteraf M. A., Barney JB. Unraveling the resource-based triangle. Manag Dec Econ 2003; 24(4): 309-323

Porter M. A. (2010). What is value in healthcare? 363(26), 2477-2481.

Rajput, S., & Singh, S. P. (2019). Identifying Industry 4.0 IoT enablers. Management Decision, 57(8), 1784-1817. <https://doi.org/10.1016/j.cie.2022.108815>

- Ray, A., & Chaudhuri, A. K. (2021). Smart healthcare disease diagnosis and patient management: Innovation, improvement and skill development. *Machine Learning with Applications*, 3, 1-24. <https://doi.org/10.1016/j.mlwa.2020.100011>
- Riahi, Y., Saikouk, T., Gunasekaran, A., & Badraoui, I. (2021). Artificial intelligence applications in supply chain: A descriptive bibliometric analysis and future research directions. *Expert Systems with Applications*, 173. <https://doi.org/10.1016/j.cie.2022.108815>
- Saberi, S., Kouhizadeh, M., & Sarkis, J. (2019 a). Blockchains and the supply chain: Findings from a broad study of practitioners. *IEEE Engineering Management Review*, 47(3), 95–103. <https://doi.org/10.1109/EMR.2019.2928264>
- Samuel Yousefi, B. M. (2022, April). An analytical approach for evaluating the impact of blockchain technology on sustainable supply chain performance. *International Journal of Production Economics*. <https://doi.org/10.1016/j.ijpe.2022.108429>
- Schuetz, S. W. and Venkatesh, V. (2020). The Rise of Human Machines: How Cognitive Computing Systems Challenge Assumptions of User-System. 460-482. <http://dx.doi.org/10.17705/1jais.00608>
- Scott, I. A. (2021). Demystifying machine learning: a primer for physicians. *Internal Medicine Journal*, 51(9), 1388–1400. <https://doi.org/10.1111/imj.15200>
- Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *J Data Warehouse*, 5.
- Silva, G., Pereira, L., Carvalho, J., Vinhas da Silva, R., & Simões, A. (2024). The competitiveness of Portugal: views from the market.
- SupplyChainBrain. (2024, January). AI Is Helping Protect Healthcare Supply Chains, Study Shows.
- Terras, H. M. B. (2017). *Gestão de Bloco Operatório*. Dissertação de mestrado, Universidade da Beira Interior. 5951_11892.pdf (ubi.pt)

Tetteh, J., Kogi, R., Yawson, A. O., Mensah, G., Biritwum, R., & Yawson, A. E. (2019). Effect of self-rated health status on functioning difficulties among older adults in Ghana: Coarsened exact matching method of analysis of the World Health Organization's study on global aging and adult health, 14(11). <https://doi.org/10.1371/JOURNAL.PONE.0224327>

Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., & Aeini, S. (2021). Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas. In *Mathematical Problems in Engineering*. Hindawi Limited. <https://doi.org/10.1155/2021/1476043>

VanVactor, J. (2011). Cognizant healthcare logistics management: ensuring resilience during crisis. Emerald Group Publishing Limited, 245-255. <http://dx.doi.org/10.1108/17595901111167114>

Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Qi Dong, J., Fabian, N., & Haenlein, M. (2021). Digital transformation: A multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889–901. <https://doi.org/10.1016/J.JBUSRES.2019.09.022>

Verhoef, P. C., & Sloot, L. M. (2006). Out-of-Stock: Reactions, antecedents, management solutions, and a future perspective. *Retailing in 21st century: Current and future trends*, 3, 239-253. http://dx.doi.org/10.1007/978-3-540-72003-4_18

Whicher, D., & Rapp, T. (2022). The Value of Artificial Intelligence for Healthcare Decision Making-Lessons Learned. <https://doi.org/10.1016/j.jval.2021.12.009>

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/J.JSIS.2019.01.003>

World Health Organization (WHO) www.worldheathorganization.com, acedido em 2024

Anexos. Jupyter Notebook


```
In [1]: #!pip install wordcloud
from wordcloud import WordCloud
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from pprint import pprint

In [2]: dados_niveis_mrp=pd.read_excel("cuf_niveis_mrp.xlsx")
        dados_indice_ruturas=pd.read_excel("cuf_indice_ruturas.xlsx")
        dados_entradas=pd.read_excel("cuf_entradas.xlsx")

In [3]: dados_indice_ruturas=pd.read_excel("cuf_indice_ruturas.xlsx")

In [4]: dados_entradas=pd.read_excel("cuf_entradas.xlsx")
```

MRP

```
In [5]: dados_niveis_mrp.head()

Out[5]:
```

	Modif.em	Material	Área MRP	Cen.	GMRP	Ptreabast	Estmáximo	SuE	DpSuE	HoraMod.
0	14.02.2023	110006869	HCIS_1044	HCIS	4002.0	4	6	10.0	NaN	12:56:11
1	13.03.2023	110006810	HCIS_1044	HCIS	4002.0	0	0	10.0	NaN	16:40:43
2	17.05.2023	110035785	HCIS_1044	HCIS	4002.0	3	4	10.0	NaN	18:02:24
3	28.06.2023	110000021	HCIS_1044	HCIS	4002.0	100	200	10.0	1044.0	14:36:11
4	28.06.2023	110000342	HCIS_1044	HCIS	4002.0	30	40	10.0	1044.0	14:34:44

```
In [6]: dados_niveis_mrp.shape

Out[6]: (568, 10)

In [7]: dados_niveis_mrp.columns

Out[7]: Index(['Modif.em', 'Material', 'Área MRP', 'Cen.', 'GMRP', 'Ptreabast', 'Estmáximo', 'SuE', 'DpSuE', 'HoraMod.'], dtype='object')
```

Ruturas

```
In [8]: dados_indice_ruturas=pd.read_excel("cuf_indice_ruturas.xlsx")

In [9]: dados_indice_ruturas = dados_indice_ruturas.transpose()
        dados_indice_ruturas.head()

Out[9]:
```

	0	1	2	3
Unnamed: 0	CUF	CL	HCIS	HCIS_1044
01/fev.	0.052	5.86	0.09	9.33
02/Feb	0.055	5.57	0.095	7.65
03/Feb	0.045	5.3	0.089	5.78
04/Feb	0.048	5.3	0.091	5.78

```
In [10]: # turn the first row into the new header
new_header = dados_indice_ruturas.iloc[0] #grab the first row for the header
dados_indice_ruturas = dados_indice_ruturas[1:] #take the data less the header row
dados_indice_ruturas.columns = new_header #set the header row as the df header

In [11]: dados_indice_ruturas.head()

Out[11]:
```

Unnamed: 0	CUF	CL	HCIS	HCIS_1044
01/fev.	0.052	5.86	0.09	9.33
02/Feb	0.055	5.57	0.095	7.65
03/Feb	0.045	5.3	0.089	5.78
04/Feb	0.048	5.3	0.091	5.78
05/Feb	0.055	5.3	0.103	5.78

```
In [12]: dados_indice_ruturas.shape

Out[12]: (469, 4)

In [13]: dados_indice_ruturas.columns

Out[13]: Index(['CUF', 'CL', 'HCIS', 'HCIS_1044'], dtype='object', name='Unnamed: 0')
```

Entradas

```
In [14]: dados_entradas.head()

Out[14]:
```

	Material	Texto breve material	Cen.	Dep.	Doc.mat.	Itm	Dt.lçto.	Qtd	UMR	Data doc.	...	Entrado em	Hora	QtdUMP	QtdUPP	Quantidade	UMB	UMP	UPP	Cen.1	Fornecedor
0	110000038	SERINGA DUAS PEÇAS LUER 5 ML	HCIS	1044	5000025851	10	01.02.2023	800.0	UND	01.02.2023	...	01.02.2023	09:31:02	800.0	800.0	800.0	UND	UND	UND	HCIS	25000194
1	110000089	PENSO NÃO TECIDO 9 X 10 CM	HCIS	1044	5000025851	5	01.02.2023	200.0	UND	01.02.2023	...	01.02.2023	09:31:02	200.0	200.0	200.0	UND	UND	UND	HCIS	25000194
2	110000261	DRENO ASPIRA RESERV 1ENTRADA S/VACUOCH14	HCIS	1044	5000025851	6	01.02.2023	20.0	UND	01.02.2023	...	01.02.2023	09:31:02	20.0	20.0	20.0	UND	UND	UND	HCIS	25000194
3	110000347	CAMPO CIRURGICO COM ADESIVO 75 X 75 CM	HCIS	1044	5000025851	7	01.02.2023	140.0	UND	01.02.2023	...	01.02.2023	09:31:02	140.0	140.0	140.0	UND	UND	UND	HCIS	25000194
4	110000437	SERINGA TRES PEÇAS PERFUSÃO OPACA 50ML*	HCIS	1044	5000025851	8	01.02.2023	15.0	UND	01.02.2023	...	01.02.2023	09:31:02	15.0	15.0	15.0	UND	UND	UND	HCIS	25000194

5 rows × 22 columns

```
In [15]: dados_entradas.shape

Out[15]: (12266, 22)
```

Data Cleaning

Remoção de colunas que não vão ser utilizadas níveis MRP

```
In [16]: columns_drop=['SuE','DpSuE','Modif.em','HoraMod.','Cen.']
dados_niveis_mrp.drop(columns_drop, axis=1, inplace=True)
print(*dados_niveis_mrp.columns.tolist(), sep='\n')
print(dados_niveis_mrp.shape)
```

```
Material
Área MRP
GMRP
Ptreatabast
Estmáximo
(560, 5)
```

Remoção de colunas que não vão ser utilizadas entradas

```
In [17]: columns_drop=['Dt.lcto.','Qtyd','UMR','Itm','Cen..1','Data.doc.']
dados_entradas.drop(columns_drop, axis=1, inplace=True)
print(*dados_entradas.columns.tolist(), sep='\n')
print(dados_entradas.shape)
```

```
Material
Texto breve material
Cen.
Dep.
Doc.mat.
Pedido
Item
Entrado em
Hora
QtyUMP
QtyUPP
Quantidade
UMR
UMP
UPP
Fornecedor
(12266, 16)
```

Remoção de materiais com colunas nulas níveis mrp

```
In [18]: print(dados_niveis_mrp.isnull().sum())
```

```
Material      0
Área MRP      0
GMRP         14
Ptreatabast   0
Estmáximo     0
dtype: int64
```

```
In [19]: print(dados_niveis_mrp.isnull().sum(axis=1))
```

```
0      0
1      0
2      0
3      0
4      0
..
555    0
556    0
557    0
558    0
559    0
Length: 560, dtype: int64
```

```
In [20]: dados_niveis_mrp.dropna(inplace=True)
dados_niveis_mrp.drop_duplicates(inplace=True)
```

```
In [21]: print(dados_niveis_mrp.isnull().sum(axis=1))
```

```
0      0
1      0
2      0
3      0
4      0
..
555    0
556    0
557    0
558    0
559    0
Length: 546, dtype: int64
```

```
In [20]: dados_niveis_mrp.dropna(inplace=True)
dados_niveis_mrp.drop_duplicates(inplace=True)
```

```
In [21]: print(dados_niveis_mrp.isnull().sum(axis=1))
```

```
0      0
1      0
2      0
3      0
4      0
..
555    0
556    0
557    0
558    0
559    0
Length: 546, dtype: int64
```

```
In [22]: print(dados_niveis_mrp.isnull().sum())
```

```
Material      0
Área MRP      0
GMRP          0
Ptreatabast   0
Estmáximo     0
dtype: int64
```

```
In [23]: dados_niveis_mrp.shape
```

```
Out[23]: (546, 5)
```

Criação de novas variáveis

In [24]: `dn= dados_niveis_mrp`

In [25]: `dn.rename(columns={'Material': 'cod_sap', 'Área MRP': 'area_mrp', 'GMRP': 'gestao_mrp', 'Ptreatast': 'nivel_min', 'Estmáximo': 'nivel_max'}, inplace=True)`

In [26]: `print(*dn.columns.tolist(), sep='\n')`

```
cod_sap
area_mrp
gestao_mrp
nivel_min
nivel_max
```

In [27]: `dn.head()`

```
Out[27]:
```

	cod_sap	area_mrp	gestao_mrp	nivel_min	nivel_max
0	110006869	HCIS_1044	4002.0	4	6
1	110006810	HCIS_1044	4002.0	0	0
2	110035785	HCIS_1044	4002.0	3	4
3	110000021	HCIS_1044	4002.0	100	200
4	110000342	HCIS_1044	4002.0	30	40

In [28]: `dn['intervalo'] = dn['nivel_max'] - dn['nivel_min']`

```
dn.head()
```

```
Out[28]:
```

	cod_sap	area_mrp	gestao_mrp	nivel_min	nivel_max	intervalo
0	110006869	HCIS_1044	4002.0	4	6	2
1	110006810	HCIS_1044	4002.0	0	0	0
2	110035785	HCIS_1044	4002.0	3	4	1
3	110000021	HCIS_1044	4002.0	100	200	100
4	110000342	HCIS_1044	4002.0	30	40	10

In [30]: `dn['intervalo'] = dn['nivel_max'] - dn['nivel_min']`

```
# Selecionar os top 25 'cod_sap' com maior variabilidade
top_25_intervalo = dn.nlargest(25, 'intervalo')
print("Top 25 Materiais com Maior Intervalo MRP:")
top_25_intervalo.head()
```

Top 25 Materiais com Maior Intervalo MRP:

```
Out[30]:
```

	cod_sap	area_mrp	gestao_mrp	nivel_min	nivel_max	intervalo
370	110006162	HCIS_1044	4002.0	3500	6000	2500
501	110003923	HCIS_1044	4002.0	3000	5000	2000
504	110006170	HCIS_1044	4002.0	1920	3000	1080
474	110000038	HCIS_1044	4002.0	1200	2000	800
317	110002683	HCIS_1044	4002.0	600	1200	600

In [31]: `# Selecionar itens com alta y acima da média
items_intervalo_alto = dn[dn['intervalo'] > dn['intervalo'].mean()]
print("Materiais com Alto Intervalo Acima da Média:")
items_intervalo_alto.head()`

Materiais com Alto Intervalo Acima da Média:

```
Out[31]:
```

	cod_sap	area_mrp	gestao_mrp	nivel_min	nivel_max	intervalo
3	110000021	HCIS_1044	4002.0	100	200	100
34	110007350	HCIS_1044	4002.0	50	100	50
55	110052505	HCIS_1044	4002.0	150	300	150
79	110000230	HCIS_1044	4002.0	250	400	150
150	110002638	HCIS_1044	4002.0	300	350	50

Dados entradas

Criação de novas variáveis

In [32]: `de= dados_entradas
print(*de.columns.tolist(), sep='\n')`

```
Material
Texto breve material
Cen.
Dep.
Doc.mat.
Pedido
Item
Entrado em
Hora
QtdUHP
QtdUPP
Quantidade
UPB
UPP
Fornecedor
```

In [33]: `de.rename(columns={'Material': 'cod_sap', 'Texto breve material': 'nome_material', 'Cen.': 'unidade_hosp', 'Doc.mat.': 'nº_encomenda', 'Pedido': 'pedido_compra', 'Item': 'qt_entrada_bci', 'Entrado em': 'data_entrada', 'Dep.': 'serviço`

In [34]: `de['data_entrada'] = pd.to_datetime(de['data_entrada'], dayfirst=True)
de['year_entrada'] = de['data_entrada'].dt.year
de['month_entrada'] = de['data_entrada'].dt.month
de['day_entrada'] = de['data_entrada'].dt.day`

```
In [35]: print(*de.columns.tolist(), sep='\n')
```

```
cod_sap
nome_material
unidade_hosp
servico_bci
nº_encomenda
pedido_compra
qt_entrada_bci
data_entrada
hora_entrada
QtdUMP
QtdUPP
Quantidade
UMB
UMP
UPP
fornecedor_hosp
year_entrada
month_entrada
day_entrada
```

```
In [36]: de.head(1)
```

```
Out[36]:
```

	cod_sap	nome_material	unidade_hosp	servico_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtdUMP	QtdUPP	Quantidade	UMB	UMP	UPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada
0	110000038	SERINGA DUAS PEÇAS LUER 5 ML	HCIS	1044	5000025851	7400326545	160	2023-02-01	09:31:02	800.0	800.0	800.0	UND	UND	UND	25000194	2023	2	1

```
In [37]: print(dados_entradas.isnull().sum())
```

```
cod_sap          0
nome_material    0
unidade_hosp     0
servico_bci      0
nº_encomenda     0
pedido_compra    0
qt_entrada_bci   0
data_entrada     0
hora_entrada     0
QtdUMP           0
QtdUPP           0
Quantidade       0
UMB              0
UMP              0
UPP              0
fornecedor_hosp  0
year_entrada     0
month_entrada    0
day_entrada      0
dtype: int64
```

```
In [38]: de.dtypes
```

```
Out[38]:
```

cod_sap	int64
nome_material	object
unidade_hosp	object
servico_bci	int64
nº_encomenda	int64
pedido_compra	int64
qt_entrada_bci	int64
data_entrada	datetime64[ns]
hora_entrada	object
QtdUMP	float64
QtdUPP	float64
Quantidade	float64
UMB	object
UMP	object
UPP	object
fornecedor_hosp	int64
year_entrada	int64
month_entrada	int64
day_entrada	int64
dtype:	object

```
In [39]: dados_entradas.dropna(inplace=True)
dados_entradas.drop_duplicates(inplace=True)
```

```
In [40]: print(dados_entradas.isnull().sum(axis=1))
```

```
0      0
1      0
2      0
3      0
4      0
..
12261   0
12262   0
12263   0
12264   0
12265   0
Length: 12266, dtype: int64
```

```
In [41]: print(dados_entradas.isnull().sum())
```

```
cod_sap          0
nome_material    0
unidade_hosp     0
servico_bci      0
nº_encomenda     0
pedido_compra    0
qt_entrada_bci   0
data_entrada     0
hora_entrada     0
QtdUMP           0
QtdUPP           0
Quantidade       0
UMB              0
UMP              0
UPP              0
fornecedor_hosp  0
year_entrada     0
month_entrada    0
day_entrada      0
dtype: int64
```

```
In [42]: de.shape
```

```
Out[42]: (12266, 19)
```

```
In [43]: de.head()
```

	cod_sap	nome_material	unidade_hosp	serviço_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtyUMP	QtyUPP	Quantidade	UMB	UMP	UPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada
0	110000038	SERINGA DUAS PEÇAS LUER 5 ML	HCIS	1044	5000025851	7400326545	160	2023-02-01	09:31:02	800.0	800.0	800.0	UND	UND	UND	25000194	2023	2	1
1	110000089	PENSO NÃO TECIDO 9 X 10 CM	HCIS	1044	5000025851	7400326545	100	2023-02-01	09:31:02	200.0	200.0	200.0	UND	UND	UND	25000194	2023	2	1
2	110000261	DRENO ASPIRA RESERV 1ENTRADA S/VACUOCH14	HCIS	1044	5000025851	7400326545	120	2023-02-01	09:31:02	20.0	20.0	20.0	UND	UND	UND	25000194	2023	2	1
3	110000347	CAMPO CIRURGICO COM ADESIVO 75 X 75 CM	HCIS	1044	5000025851	7400326545	130	2023-02-01	09:31:02	140.0	140.0	140.0	UND	UND	UND	25000194	2023	2	1
4	110000437	SERINGA TRES PEÇAS PERFUSÃO OPACA 50ML*	HCIS	1044	5000025851	7400326545	140	2023-02-01	09:31:02	15.0	15.0	15.0	UND	UND	UND	25000194	2023	2	1

```
In [44]: de.dtypes
```

Out[44]:	cod_sap	int64
	nome_material	object
	unidade_hosp	object
	serviço_bci	int64
	nº_encomenda	int64
	pedido_compra	int64
	qt_entrada_bci	int64
	data_entrada	datetime64[ns]
	hora_entrada	object
	QtyUMP	float64
	QtyUPP	float64
	Quantidade	float64
	UMB	object
	UMP	object
	UPP	object
	fornecedor_hosp	int64
	year_entrada	int64
	month_entrada	int64
	day_entrada	int64
	dtype:	object

```
In [45]: nome_material_counts = de['nome_material'].value_counts()
print(nome_material_counts)
```

```
ELECTRODO BIS ADULTO      105
TROUXA UNIVERSAL          99
TROUXA EXTREMIDADES       97
SERINGA DUAS PEÇAS LUER 20 ML  96
SERINGA DUAS PEÇAS LUER 5 ML  93
...
TUBO ENDOTRAQUEAL SEM BALÃO 3,0MM  1
INTRODUTOR AGULHA RAQUI 20G X 35 MM  1
LINHA DE AMOSTRA P/VENTILADOR  1
TUBO SILICONE TRANSLUCIDO ROLO 5X8MM  1
MASCARA LARINGEA ARAMADA Nº3  1
Name: nome_material, Length: 520, dtype: int64
```

```
In [46]: cod_sap_counts = de['cod_sap'].value_counts()
print(cod_sap_counts)
```

```
110004699      105
110001177       99
110002781       97
110000061       96
110000038       93
...
110002879        1
110002864        1
110001140        1
110004913        1
110021200        1
Name: cod_sap, Length: 525, dtype: int64
```

```
In [47]: cod_sap_more_than_once = cod_sap_counts[cod_sap_counts > 1]
num_cod_sap_more_than_once = len(cod_sap_more_than_once)
print(f"número de materiais associados ao 'cod_sap' mais de uma vez: {num_cod_sap_more_than_once}")

número de materiais associados ao 'cod_sap' mais de uma vez: 474
```

```
In [48]: pedido_compra_counts = de['pedido_compra'].value_counts()
print(pedido_compra_counts)
```

```
7400326133      150
7400335137      105
7400344191      100
7400338671      100
7400331379       83
...
7400334565        1
7400344028        1
7400338316        1
7400338426        1
7400336172        1
Name: pedido_compra, Length: 603, dtype: int64
```

```
In [51]: pedido_compra_counts = de['pedido_compra'].value_counts()
cod_sap_counts = de['cod_sap'].value_counts()

print(f"Total de linhas contadas em 'pedido_compra': {pedido_compra_counts.sum()}")
print(f"Total de linhas contadas em 'cod_sap': {cod_sap_counts.sum()}")

Total de linhas contadas em 'pedido_compra': 12266
Total de linhas contadas em 'cod_sap': 12266
```

```
In [54]: materiais_rep_mais_50 = count_materiais_rep[count_materiais_rep > 50]
```

```
print("materiais repetidos mais de 50 vezes:")
print(materiais_rep_mais_50)
```

```
materiais repetidos mais de 50 vezes:
ELECTRODO BIS ADULTO      105
TROUXA UNIVERSAL          99
TROUXA EXTREIMIDADES      97
SERINGA DUAS PEÇAS LUER 20 ML  96
SERINGA DUAS PEÇAS LUER 5 ML  93
...
CATETER ENDOVE PERIFERICO PROTEC14GX45MM  52
LIGADURA ESTERIL CREPE 4MX10CM          52
OCULO NASAL ADULTO          51
TUBO GUEDELL N°3           51
LUVA CIRURGICA SEH PO N° 8,5             51
Name: nome_material, Length: 70, dtype: int64
```

```
In [55]: encomendas_rep = dados_cod_sap_rep['nº_encomenda'].value_counts()
print("encomendas rep:")
print(encomendas_rep)
```

```
encomendas rep:
5000030332    95
5000203042    85
5000192733    74
5000151355    74
5000094553    74
..
5000095461     1
5000095439     1
5000095451     1
5000152545     1
5000206015     1
Name: nº_encomenda, Length: 1743, dtype: int64
```

Analisar Distribuição ao Longo do Tempo

```
In [57]: de['data_entrada'] = pd.to_datetime(de['data_entrada'])
```

```
In [58]: encomendas_dia = de.groupby('data_entrada')['nº_encomenda'].count()
top_5_dias_encomendas = encomendas_dia.sort_values(ascending=False).head(5)
```

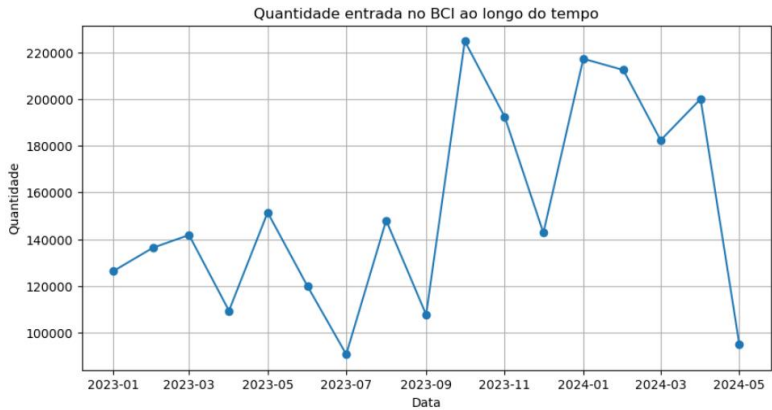
```
print("Top 5 dias com mais encomendas:")
print(top_5_dias_encomendas)

Top 5 dias com mais encomendas:
data_entrada
2023-04-11    102
2023-02-08    101
2024-02-13     97
2024-02-12     93
2024-04-09     92
Name: nº_encomenda, dtype: int64
```

```
In [60]: de['data_entrada'] = pd.to_datetime(de['data_entrada'], yearfirst=True)
```

```
# Agrupar por mês e somar as quantidades de entrada
monthly_data = de.groupby(de['data_entrada'].dt.to_period('M'))['qt_entrada_bci'].sum().reset_index()
monthly_data['data_entrada'] = monthly_data['data_entrada'].dt.to_timestamp()

# Plotar a tendência das quantidades de entrada ao longo do tempo
plt.figure(figsize=(10, 5))
plt.plot(monthly_data['data_entrada'], monthly_data['qt_entrada_bci'], marker='o')
plt.title('Quantidade entrada no BCI ao longo do tempo')
plt.xlabel('Data')
plt.ylabel('Quantidade')
plt.grid(True)
plt.show()
```



```
In [61]: pedidos_dia = de.groupby('data_entrada')['qt_entrada_bci'].count()
top_5_dias_qt_entrada_bci = pedidos_dia.sort_values(ascending=False).head(5)
```

```
print("Top 5 dias com qt_entrada_bci:")
print(top_5_dias_qt_entrada_bci)

Top 5 dias com qt_entrada_bci:
data_entrada
2023-04-11    102
2023-02-08    101
2024-02-13     97
2024-02-12     93
2024-04-09     92
Name: qt_entrada_bci, dtype: int64
```

```
In [62]: mais_pedidos = de.groupby('cod_sap')['data_entrada'].count()
top_5_mais_pedidos = mais_pedidos.sort_values(ascending=False).head(5)
```

```
print("Top 5 materiais mais pedidos:")
print(top_5_mais_pedidos)

Top 5 materiais mais pedidos:
cod_sap
110004699    105
110001177     99
110002781     97
110000061     96
110000038     93
Name: data_entrada, dtype: int64
```

```
In [63]: # Classificar materiais em classes A, B e C
def classificar_abc(percentual):
    if percentual <= 70:
        return 'A'
    elif percentual <= 90:
        return 'B'
    else:
        return 'C'

# Supondo que 'de' é o DataFrame original com os dados
de['consumo_total'] = de['qt_entrada_bci']

# Ordenar materiais pelo consumo total em ordem decrescente
de.sort_values(by='consumo_total', ascending=False, inplace=True)

# Calcular o consumo acumulado e a porcentagem do consumo total
de['consumo_acumulado'] = de['consumo_total'].cumsum()
de['percentual_consumo_total'] = 100 * de['consumo_acumulado'] / de['consumo_total'].sum()

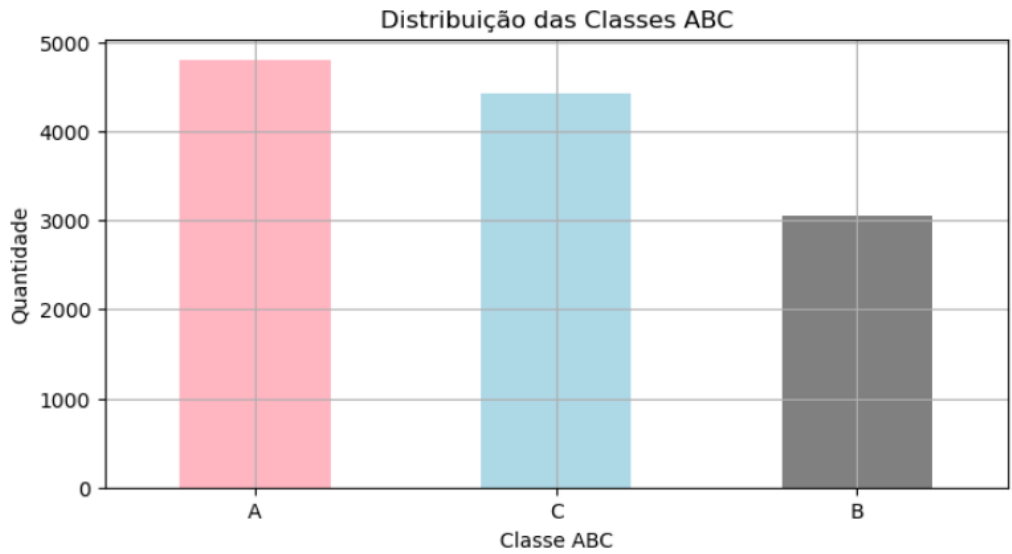
# Aplicar a classificação ABC
de['classe_abc'] = de['percentual_consumo_total'].apply(classificar_abc)

# Exibir o DataFrame com as classificações ABC
print(de[['consumo_total', 'consumo_acumulado', 'percentual_consumo_total', 'classe_abc']].head())

# Contar quantas vezes cada classe ocorre
abc_counts = de['classe_abc'].value_counts()

# Plotar o gráfico de barras
plt.figure(figsize=(8, 4))
# Especificar as cores como rosa claro, azul claro e cinzento
colors = ['#FFB6C1', '#ADD8E6', '#808080'] # Rosa claro, Azul claro, Cinzento
abc_counts.plot(kind='bar', color=colors)
plt.title('Distribuição das Classes ABC')
plt.xlabel('Classe ABC')
plt.ylabel('Quantidade')
plt.xticks(rotation=0) # Manter os rótulos das classes sem rotação
plt.grid(True)
plt.show()
```

	consumo_total	consumo_acumulado	percentual_consumo_total	classe_abc
6845	1560	1560	0.060028	A
1001	1560	3120	0.120057	A
1003	1510	4630	0.178161	A
992	1460	6090	0.234342	A
991	1450	7540	0.290137	A



```
In [65]: top_5_A = top_5_materials_by_class[top_5_materials_by_class['classe_abc'] == 'A']
top_5_B = top_5_materials_by_class[top_5_materials_by_class['classe_abc'] == 'B']
top_5_C = top_5_materials_by_class[top_5_materials_by_class['classe_abc'] == 'C']
top_5_A.head()

Out[65]:
```

	cod_sap	nome_material	unidade_hosp	serviço_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtDUMP	...	UMP	UPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada	consumo_total	consumo_acumulado	percentual_consumo_total	classe_abc
0	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA	HCIS	1044	5000157605	7400335137	1560	2023-08-17	16:50:39	75.0	...	UND	UND	25000194	2023	8	17	1560	1560	0.060028	A
1	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA	HCIS	1044	5000151355	7400335137	1560	2023-08-03	11:25:54	75.0	...	UND	UND	25000194	2023	8	3	1560	3120	0.120057	A
2	110000746	SISTEMA INFUSÃO FLUIDOS BIPERFURADOR	HCIS	1044	5000151355	7400335137	1510	2023-08-03	11:25:54	25.0	...	UND	UND	25000194	2023	8	3	1510	4630	0.178161	A
3	110000369	MASCARA OXIGENIO ADULTO	HCIS	1044	5000151355	7400335137	1460	2023-08-03	11:25:54	17.0	...	UND	UND	25000194	2023	8	3	1460	6090	0.234342	A
4	110000321	OCULO NASAL ADULTO	HCIS	1044	5000151355	7400335137	1450	2023-08-03	11:25:54	50.0	...	UND	UND	25000194	2023	8	3	1450	7540	0.290137	A

5 rows × 23 columns

In [66]:

top_5_B.head()

Out[66]:

	cod_sap	nome_material	unidade_hosp	serviço_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtyUMP	...	UMP	UPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada	consumo_total	consumo_acumulado	percentual_consumo_total	classe_abc
5	110000118	PENSO NÃO TECIDO 9 X 15 CM	HCIS	1044	5000177925	7400336086	220	2023-09-20	08:49:37	100.0	...	UND	UND	25000194	2023	9	20	220	1819190	70.001962	B
6	110026349	OBTURADOR BIDIRECIONAL	HCIS	1044	5000085672	7400330554	220	2023-04-27	12:24:16	100.0	...	UND	UND	25000194	2023	4	27	220	1848010	71.110949	B
7	110006885	MASCARA LARINGEA I-GEL Nº3.0	HCIS	1044	5000062883	7400346417	220	2024-03-27	15:59:14	24.0	...	UND	UND	25000194	2024	3	27	220	1846030	71.034759	B
8	110002997	BATA CIRURGICA REFORÇADA TAM XL	HCIS	1044	5000024081	7400343673	220	2024-02-02	16:40:15	96.0	...	UND	UND	25000194	2024	2	2	220	1846250	71.043224	B
9	110003616	SUPORTE BRAÇO ADULTO	HCIS	1044	5000244740	7400341048	220	2023-12-22	13:58:12	10.0	...	UND	UND	25000194	2023	12	22	220	1846470	71.051690	B

5 rows × 23 columns

In [67]:

top_5_C.head()

Out[67]:

	cod_sap	nome_material	unidade_hosp	serviço_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtyUMP	...	UMP	UPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada	consumo_total	consumo_acumulado	percentual_consumo_total	classe_abc
10	110000971	LIGADURA ELASTICA FORTE COMPRESSÃO 10 CM	HCIS	1044	5000169790	7400336357	120	2023-09-06	14:41:30	30.0	...	UND	UND	25000194	2023	9	6	120	2338900	90.000269	C
11	110000012	PENSO RÁPIDO ADULTO	HCIS	1044	5000192733	7400337945	120	2023-10-11	17:13:26	100.0	...	UND	UND	25000194	2023	10	11	120	2357140	90.702140	C
12	110000499	CATETER ENDOVE PERIFÉRICO PROTEC14GX45MM	HCIS	1044	5000181561	7400337189	120	2023-09-25	16:20:04	50.0	...	UND	UND	25000194	2023	9	25	120	2355820	90.651347	C
13	110004555	LIGADURA ELASTICA AUTO-ADERENTE 15 CM	HCIS	1044	5000085677	7400330850	120	2023-04-27	12:26:58	5.0	...	UND	UND	25000194	2023	4	27	120	2355940	90.655964	C
14	110000680	REFERENCIA VASCULAR VERMELHA	HCIS	1044	5000067304	7400329476	120	2023-03-29	11:47:46	2.0	...	CX	CX	25000194	2023	3	29	120	2356060	90.660582	C

5 rows × 23 columns

Dados Ruturas

In [68]:

dr = dados_indice_ruturas

In [69]:

dr.rename(columns={'Unnamed: 0': 'data'},inplace=True)

In [70]:

dr.head()

Out[70]:

Unnamed: 0	CUF	CL	HCIS	HCIS_1044
01/fev	0.052	5.86	0.09	9.33
02/fev	0.055	5.57	0.095	7.65
03/fev	0.045	5.3	0.089	5.78
04/fev	0.048	5.3	0.091	5.78
05/fev	0.055	5.3	0.103	5.78

In [71]:

dr = dr.dropna()
dr.isna().sum()

Out[71]:

Unnamed: 0
CUF
CL
HCIS
HCIS_1044
dtype: int64

0
0
0
0
0

In [72]:

dr.shape

Out[72]:

(461, 4)

In [73]:

dr.head()

Out[73]:

Unnamed: 0	CUF	CL	HCIS	HCIS_1044
01/fev	0.052	5.86	0.09	9.33
02/fev	0.055	5.57	0.095	7.65
03/fev	0.045	5.3	0.089	5.78
04/fev	0.048	5.3	0.091	5.78
05/fev	0.055	5.3	0.103	5.78

In [74]:

dr.rename({'CUF': 'cuf_geral','CL': 'centro_log','HCIS': 'cuf_tejo','HCIS_1044': 'serviço_bci'}, inplace=True, axis=1)

print("Nomes das colunas depois de renomear:", dr.columns)

Multiplique as colunas 'cuf_geral' e 'cuf_tejo' por 100 e converta para inteiros
dr['cuf_geral'] = (dr['cuf_geral'] * 100).astype(float)
dr['cuf_tejo'] = (dr['cuf_tejo'] * 100).astype(float)
#converter o tipo de object para float para se poder fazer operacoes
dr['centro_log'] = dr['centro_log'].astype(float)
dr['serviço_bci'] = dr['serviço_bci'].astype(float)

dr.head()

Out[74]:

Unnamed: 0	cuf_geral	centro_log	cuf_tejo	serviço_bci
01/fev	5.2	5.86	9.0	9.33
02/fev	5.5	5.57	9.5	7.65
03/fev	4.5	5.30	8.9	5.78
04/fev	4.8	5.30	9.1	5.78
05/fev	5.5	5.30	10.3	5.78


```
In [75]: # Mostrar o DataFrame com 'data' como índice
print(dr)

# Crie a figura do gráfico
plt.figure(figsize=(8, 4))

# Plote cada coluna do DataFrame
for i, column in enumerate(dr.columns):
    plt.plot(dr.index, dr[column], label=column, color=colors[i % len(colors)]) # Use modulo para repetir cores se necessário

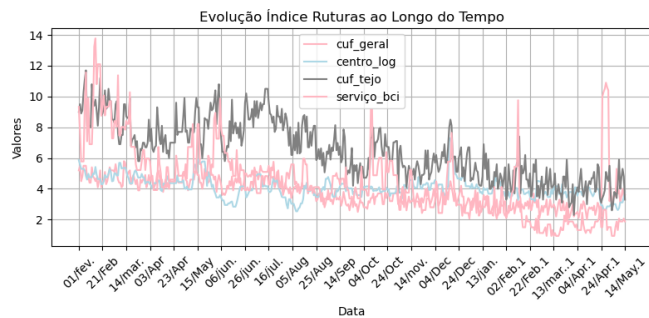
colors = ['pink', 'lightblue', 'grey', 'navy'] # Adicione mais cores se tiver mais colunas

# Títulos e Legendas
plt.title('Evolução Índice Ruturas ao Longo do Tempo')
plt.xlabel('Data')
plt.ylabel('Valores')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.tight_layout()
ax = plt.gca()
ax.set_xticks(ax.get_xticks()[::20])

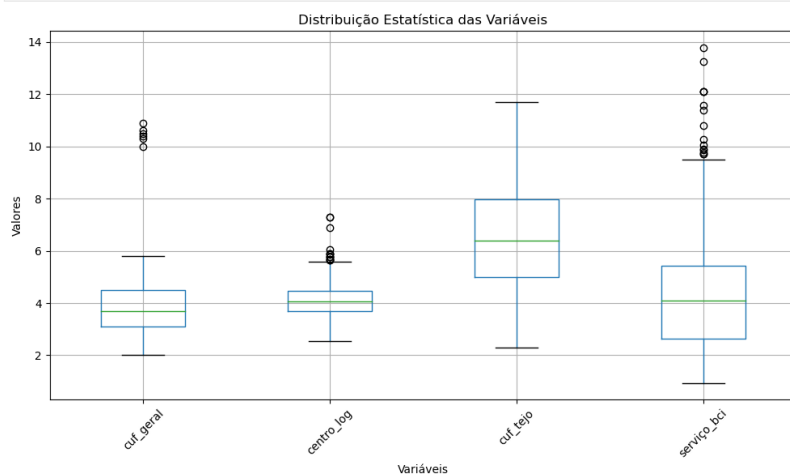
# Exiba o gráfico
plt.show()

Unnamed: 0    cuf_geral    centro_log    cuf_tejo    serviço_bci
01/fev.        5.2         5.86         9.0         9.33
02/fev.        5.5         5.57         9.5         7.65
03/fev.        4.5         5.30         8.9         5.78
04/fev.        4.8         5.30         9.1         5.78
05/fev.        5.5         5.30         10.3        5.78
...          ...          ...          ...          ...
10/May.1       3.1         3.12         4.0         1.88
11/May.1       3.5         3.12         4.7         1.88
12/May.1       4.0         3.12         5.3         1.88
13/May.1       3.5         3.43         4.8         2.07
14/May.1       3.6         3.78         3.3         1.88

[461 rows x 4 columns]
```



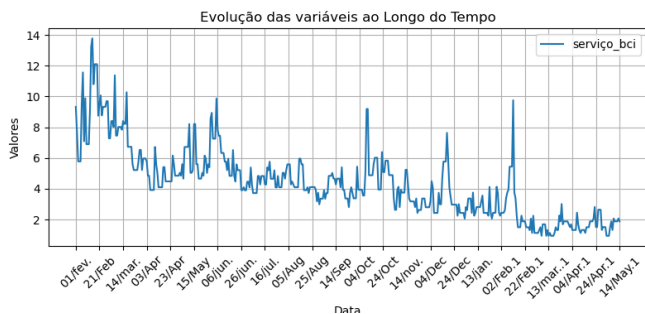
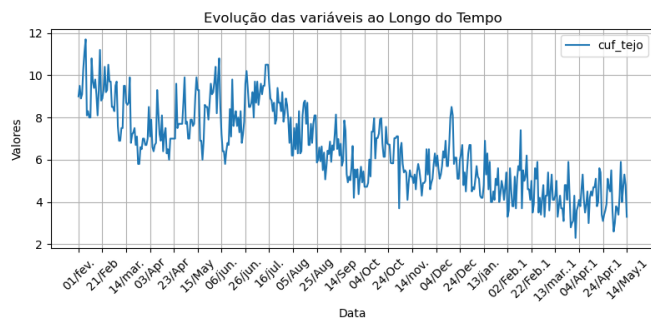
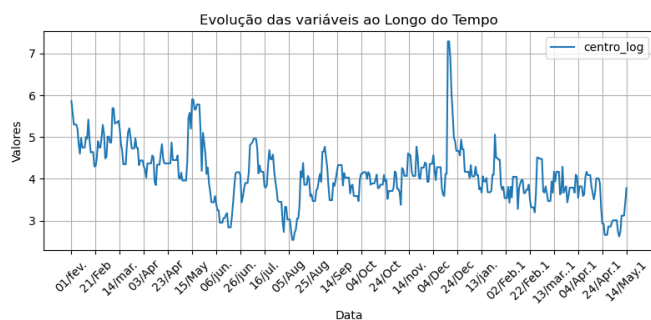
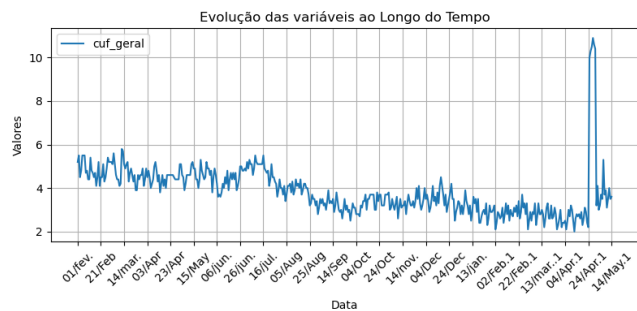
```
In [77]: plt.figure(figsize=(10, 6))
dr.boxplot()
plt.title('Distribuição Estatística das Variáveis')
plt.xlabel('Variáveis')
plt.ylabel('Valores')
plt.xticks(rotation=45)
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
In [79]: for column in dr.columns:
plt.figure(figsize=(8,4))
plt.plot(dr.index, dr[column], label=column)
plt.title(f'Evolução das variáveis ao Longo do Tempo')
plt.xlabel('Data')
plt.ylabel('Valores')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.tight_layout()

ax = plt.gca()
ax.set_xticks(ax.get_xticks()[::20])

plt.show()
```



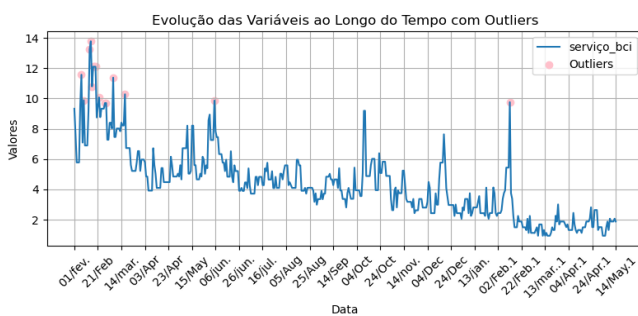
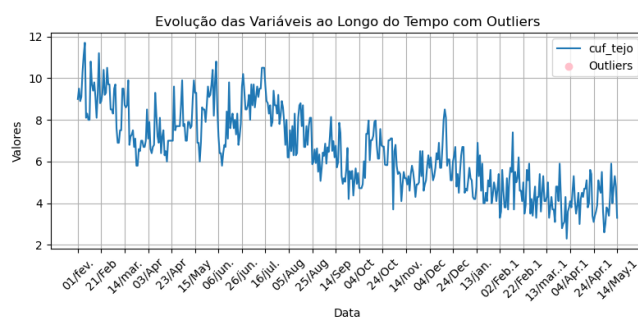
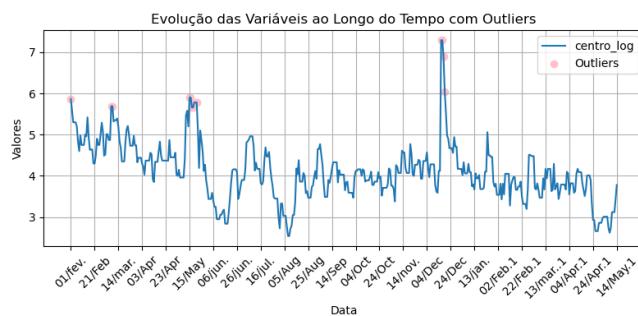
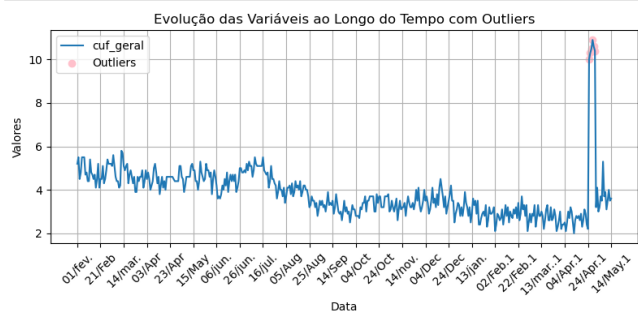
```
In [80]: # Calcular os quartis e o intervalo interquartil para cada variável
Q1 = dr.quantile(0.25)
Q3 = dr.quantile(0.75)
IQR = Q3 - Q1

# Identificar outliers para cada variável
outliers = ((dr < (Q1 - 1.5 * IQR)) | (dr > (Q3 + 1.5 * IQR)))

# Plotar os dados e destacar os outliers
for column in dr.columns:
    plt.figure(figsize=(8, 4))
    plt.plot(dr.index, dr[column], label=column)

    # Destacar outliers
    outlier_indices = outliers[outliers[column] == True].index
    plt.scatter(outlier_indices, dr.loc[outlier_indices, column], color='pink', label='Outliers')

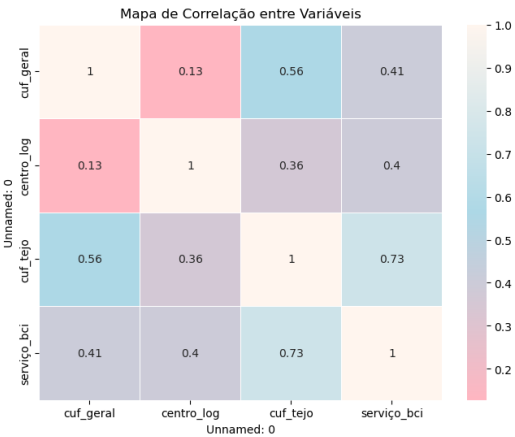
plt.title('Evolução das Variáveis ao Longo do Tempo com Outliers')
plt.xlabel('Data')
plt.ylabel('Valores')
plt.xticks(rotation=45)
plt.legend()
plt.grid(True)
plt.tight_layout()
ax = plt.gca()
ax.set_xticks(ax.get_xticks()[::20])
plt.show()
```



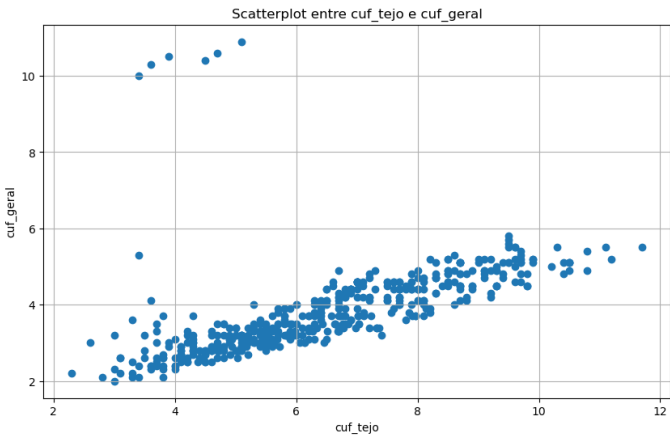
```
In [81]: import seaborn as sns
from matplotlib.colors import LinearSegmentedColormap

# Cores desejadas
colors = ["lightpink", "lightblue", "#FFF5EE"] # Rosa, Azul e Cinzento
cmap = LinearSegmentedColormap.from_list("custom_cmap", colors, N=256)

plt.figure(figsize=(8, 6))
sns.heatmap(dr.corr(), annot=True, cmap=cmap, linewidths=0.5)
plt.title('Mapa de Correlação entre Variáveis')
plt.show()
```



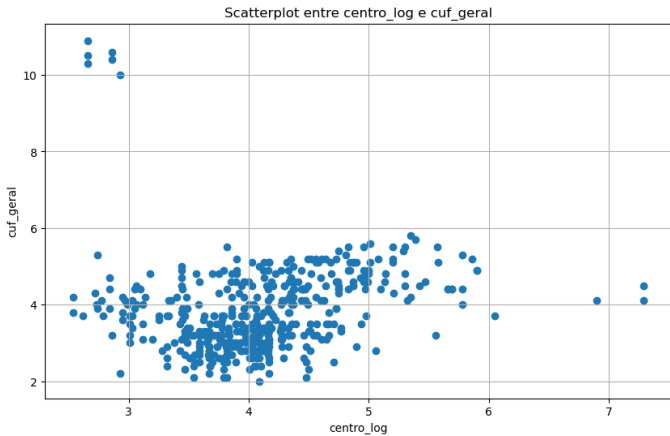
```
In [82]: plt.figure(figsize=(10, 6))
plt.scatter(dr['cuf_tejo'], dr['cuf_geral'])
plt.title('Scatterplot entre cuf_tejo e cuf_geral')
plt.xlabel('cuf_tejo')
plt.ylabel('cuf_geral')
plt.grid(True)
plt.show()
```



```
In [83]: correlation = dr['cuf_tejo'].corr(dr['cuf_geral'])
print(f'Correlação de Pearson entre cuf_tejo e cuf_geral: {correlation:.2f}')
Correlação de Pearson entre cuf_tejo e cuf_geral: 0.56
```

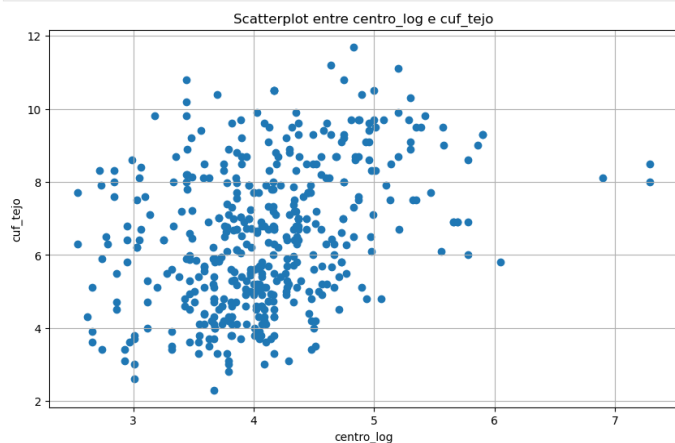
O gráfico de dispersão apresentado acima relaciona as variáveis cuf_tejo e cuf_geral através de pontos, cada um representativo de um valor. Observa-se que o aumento do índice de ruturas afeto À CUF Tejo, é sinónimo de um maior índice de ruturas afeto à CUF.

```
In [84]: plt.figure(figsize=(10, 6))
plt.scatter(dr['centro_log'], dr['cuf_geral'])
plt.title('Scatterplot entre centro_log e cuf_geral')
plt.xlabel('centro_log')
plt.ylabel('cuf_geral')
plt.grid(True)
plt.show()
```



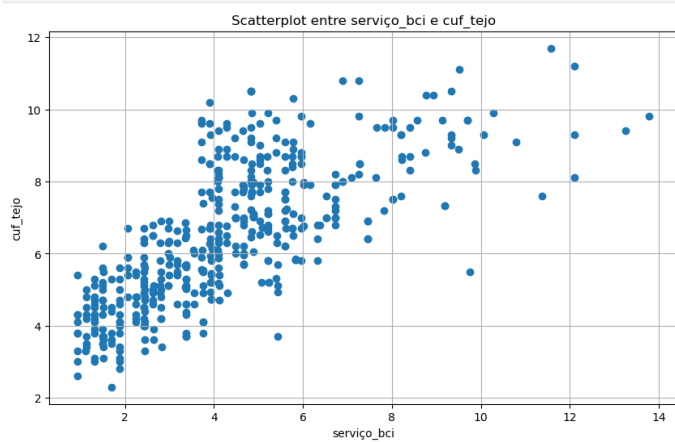
```
In [85]: correlation = dr['centro_log'].corr(dr['cuf_geral'])
print(f'Correlação de Pearson entre centro_log e cuf_geral: {correlation:.2f}')
Correlação de Pearson entre centro_log e cuf_geral: 0.13
```

```
In [86]: plt.figure(figsize=(10, 6))
plt.scatter(dr['centro_log'], dr['cuf_tejo'])
plt.title('Scatterplot entre centro_log e cuf_tejo')
plt.xlabel('centro_log')
plt.ylabel('cuf_tejo')
plt.grid(True)
plt.show()
```



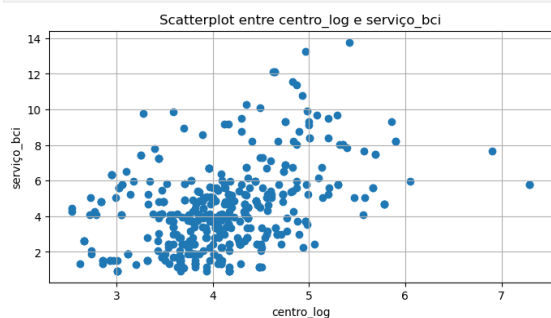
```
In [87]: correlation = dr['centro_log'].corr(dr['cuf_tejo'])
print(f'Correlação de Pearson entre centro_log e cuf_tejo: {correlation:.2f}')
Correlação de Pearson entre centro_log e cuf_tejo: 0.36
```

```
In [88]: plt.figure(figsize=(10, 6))
plt.scatter(dr['serviço_bci'], dr['cuf_tejo'])
plt.title('Scatterplot entre serviço_bci e cuf_tejo')
plt.xlabel('serviço_bci')
plt.ylabel('cuf_tejo')
plt.grid(True)
plt.show()
```

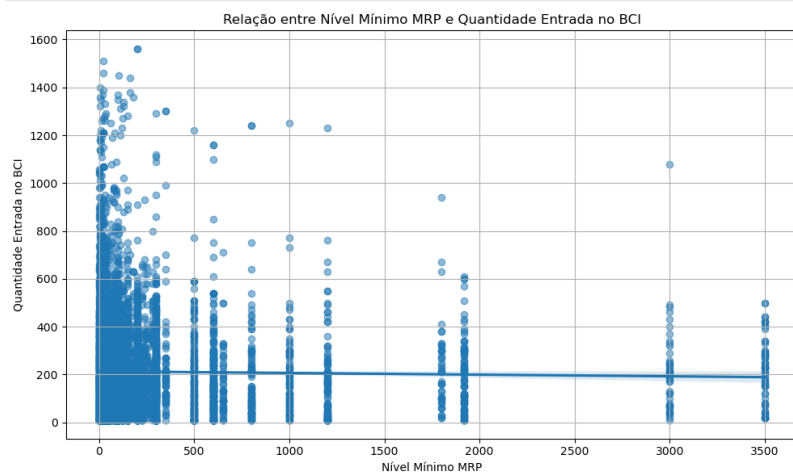


```
In [89]: correlation = dr['serviço_bci'].corr(dr['cuf_tejo'])
print(f'Correlação de Pearson entre serviço_bci e cuf_tejo: {correlation:.2f}')
Correlação de Pearson entre serviço_bci e cuf_tejo: 0.73
```

```
In [90]: plt.figure(figsize=(8, 4))
plt.scatter(dr['centro_log'], dr['serviço_bci'])
plt.title('Scatterplot entre centro_log e serviço_bci')
plt.xlabel('centro_log')
plt.ylabel('serviço_bci')
plt.grid(True)
plt.show()
```

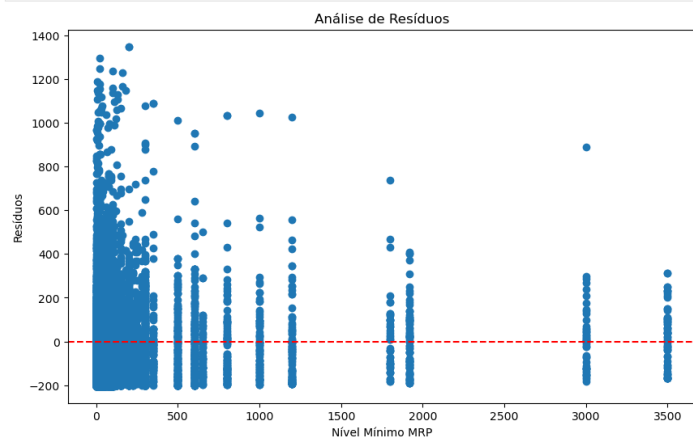


```
In [93]: #adicionar linha de regressão linear para visualizar a tendência geral
plt.figure(figsize=(10, 6))
sns.regplot(x='nivel_min', y='qt_entrada_bci', data=merged_df, scatter_kws={'alpha':0.5})
plt.title('Relação entre Nível Mínimo MRP e Quantidade Entrada no BCI')
plt.xlabel('Nível Mínimo MRP')
plt.ylabel('Quantidade Entrada no BCI')
plt.grid(True)
plt.tight_layout()
plt.show()
```



```
In [94]: import statsmodels.api as sm
# Modelo de regressão linear
X = sm.add_constant(merged_df['nivel_min']) # adicionar uma constante (intercepto)
model = sm.OLS(merged_df['qt_entrada_bci'], X).fit()

# Verificando os resíduos
residuals = model.resid
plt.figure(figsize=(10, 6))
plt.scatter(merged_df['nivel_min'], residuals)
plt.axhline(y=0, color='r', linestyle='--')
plt.title('Análise de Resíduos')
plt.xlabel('Nível Mínimo MRP')
plt.ylabel('Resíduos')
plt.show()
```



Resultados

```
In [96]: print(dados_entradas.columns)

Index(['cod_sap', 'nome_material', 'unidade_hosp', 'servico_bci',
       'nº_encomenda', 'pedido_compra', 'qt_entrada_bci', 'data_entrada',
       'hora_entrada', 'QtUUP', 'QtUUPP', 'Quantidade', 'UNB', 'UUP', 'UUPP',
       'fornecedor_hosp', 'year_entrada', 'month_entrada', 'day_entrada',
       'consumo_total', 'consumo_acumulado', 'percentual_consumo_total',
       'classe_abc'],
      dtype='object')

In [97]: dados_entradas.head()

Out[97]:
```

	cod_sap	nome_material	unidade_hosp	servico_bci	nº_encomenda	pedido_compra	qt_entrada_bci	data_entrada	hora_entrada	QtUUP	...	UUP	UUPP	fornecedor_hosp	year_entrada	month_entrada	day_entrada	consumo_total	consumo_acumulado	percentual_consumo_total	classe_abc
6845	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA	HCIS	1044	5000157605	7400335137	1560	2023-08-17	16:50:39	75.0	...	UND	UND	25000194	2023	8	17	1560	1560	0.060028	A
1001	110000583	TRAQUEIA VENTIL ADUL ANGU EXTEN S/TAMPA	HCIS	1044	5000151355	7400335137	1560	2023-08-03	11:25:54	75.0	...	UND	UND	25000194	2023	8	3	1560	3120	0.120057	A
1003	110000746	SISTEMA INFUSÃO FLUIDOS BIPERFURADOR	HCIS	1044	5000151355	7400335137	1510	2023-08-03	11:25:54	25.0	...	UND	UND	25000194	2023	8	3	1510	4630	0.178161	A
992	110000369	MASCARA OXIGENIO ADULTO	HCIS	1044	5000151355	7400335137	1460	2023-08-03	11:25:54	17.0	...	UND	UND	25000194	2023	8	3	1460	6090	0.234342	A
991	110000321	OCULO NASAL ADULTO	HCIS	1044	5000151355	7400335137	1450	2023-08-03	11:25:54	50.0	...	UND	UND	25000194	2023	8	3	1450	7540	0.290137	A

5 rows × 23 columns

```
In [142]: dados_entradas_experiment = dados_entradas.drop(['nome_material', 'unidade_hosp', 'servico_bci', 'nº_encomenda', 'pedido_compra', 'qt_entrada_bci', 'data_entrada', 'hora_entrada', 'QtUUP', 'QtUUPP', 'Quantidade', 'UNB', 'UUP', 'UUPP', 'fornecedor_hosp', 'day_entrada', 'consumo', 'percentual_consumo_total', 'classe_abc', 'qt_entrada_bci'], axis=1)
dados_entradas_experiment.head()

Out[143]:
```

	cod_sap	year_entrada	month_entrada
6845	110000583	2023	8
1001	110000583	2023	8
1003	110000746	2023	8
992	110000369	2023	8
991	110000321	2023	8

```
In [142]: grouped_total_month_encomendas = dados_entradas_experiment.groupby('year_entrada').size().reset_index(name='total_month_orders')
print(grouped_total_month_encomendas)

   year_entrada  total_month_orders
0          2023                 8289
1          2024                 4857

In [143]: dados_entradas_experiment = dados_entradas_experiment.groupby(['cod_sap', 'year_entrada', 'month_entrada']).size().reset_index(name='total_month_orders')
dados_entradas_experiment.head()

Out[143]:
```

	cod_sap	year_entrada	month_entrada	total_month_orders
0	110000012	2023	1	2
1	110000012	2023	3	1
2	110000012	2023	5	1
3	110000012	2023	6	1
4	110000012	2023	8	2

```
In [144]: max_value = dados_entradas_experiment['total_month_orders'].max()
print(max_value)

14

In [145]: average_orders_per_month = dados_entradas_experiment['total_month_orders'].mean()
print(average_orders_per_month.round(3))

2.428

In [146]: std_deviation = dados_entradas_experiment['total_month_orders'].std()
print(round(std_deviation, 3))

1.62

In [147]: dados_entradas_experiment['total_month_orders'].unique()

Out[147]: array([ 2,  1,  3,  6,  7,  4,  5,  8, 10,  9, 12, 11, 14], dtype=int64)

In [148]: # prompt: unique values for column cod_sap count
unique_total_month_encomendas = dados_entradas_experiment['cod_sap'].nunique()
print(unique_total_month_encomendas)

525

In [152]: # prompt: add rows with zero count for the missing months for cod_sap 110000012

# Create a DataFrame with all possible year-month combinations
all_year_months = pd.MultiIndex.from_product([
    dados_entradas_experiment['cod_sap'].unique(),
    [2023, 2024],
    [1,2,3,4,5,6,7,8,9,10,11,12],
], names=['cod_sap', 'year_entrada', 'month_entrada'])

# Reindex to include missing months
dados_cod_sap_filled = dados_entradas_experiment.set_index(['cod_sap', 'year_entrada', 'month_entrada']).reindex(all_year_months, fill_value=0).reset_index()

print(dados_cod_sap_filled.head())
```

	cod_sap	year_entrada	month_entrada	total_month_orders
0	110000012	2023	1	2
1	110000012	2023	2	0
2	110000012	2023	3	1
3	110000012	2023	4	0
4	110000012	2023	5	1

```
In [153]: # prompt: filter dataframe for cod_sap 110000038
dados_cod_sap_filled[dados_cod_sap_filled['cod_sap'] == 110000012]
```

Out[153]:

	cod_sap	year_entrada	month_entrada	total_month_orders
0	110000012	2023	1	2
1	110000012	2023	2	0
2	110000012	2023	3	1
3	110000012	2023	4	0
4	110000012	2023	5	1
5	110000012	2023	6	1
6	110000012	2023	7	0
7	110000012	2023	8	2
8	110000012	2023	9	1
9	110000012	2023	10	1
10	110000012	2023	11	1
11	110000012	2023	12	3
12	110000012	2024	1	2
13	110000012	2024	2	1
14	110000012	2024	3	1
15	110000012	2024	4	2
16	110000012	2024	5	0
17	110000012	2024	6	0
18	110000012	2024	7	0
19	110000012	2024	8	0
20	110000012	2024	9	0
21	110000012	2024	10	0
22	110000012	2024	11	0
23	110000012	2024	12	0

```
In [154]: dados_entradas_experiment = dados_cod_sap_filled
```

```
In [157]: # prompt: filter dataframe removing everything with year_entrada equals 2024 and month_entrada bigger than 6, inclusive
dados_entradas_experiment = dados_entradas_experiment[~((dados_entradas_experiment['year_entrada'] == 2024) & (dados_entradas_experiment['month_entrada'] >= 6))]
dados_entradas_experiment[dados_entradas_experiment['cod_sap'] == 110000012]
```

Out[157]:

	cod_sap	year_entrada	month_entrada	total_month_orders
0	110000012	2023	1	2
1	110000012	2023	2	0
2	110000012	2023	3	1
3	110000012	2023	4	0
4	110000012	2023	5	1
5	110000012	2023	6	1
6	110000012	2023	7	0
7	110000012	2023	8	2
8	110000012	2023	9	1
9	110000012	2023	10	1
10	110000012	2023	11	1
11	110000012	2023	12	3
12	110000012	2024	1	2
13	110000012	2024	2	1
14	110000012	2024	3	1
15	110000012	2024	4	2
16	110000012	2024	5	0

```
In [159]: average_orders_per_month = dados_entradas_experiment['total_month_orders'].mean()
print(average_orders_per_month.round(3))
```

```
In [160]: std_deviation = dados_entradas_experiment['total_month_orders'].std()
print(round(std_deviation, 3))
```

```
In [161]: sum = average_orders_per_month.round(3) + std_deviation
# Imprimindo o resultado
print(sum)
```

```
In [163]: dados_entradas_experiment = dados_entradas_experiment[dados_entradas_experiment['total_month_orders'] > sum]
dados_entradas_experiment.head()
```

Out[163]:

	cod_sap	year_entrada	month_entrada	total_month_orders
37	110000018	2024	2	6
60	110000020	2024	1	7
61	110000020	2024	2	4
121	110000025	2023	2	4
122	110000025	2023	3	5

```
In [164]: # Definindo o limite com base na soma
limite = average_orders_per_month.round(3) + std_deviation

# Criando a nova coluna com valores que são maiores que 'sum'
dados_entradas_experiment['max_variavel'] = dados_entradas_experiment['total_month_orders'].apply(lambda x: 1 if x > limite else 0)

# Imprimindo as primeiras linhas para verificar
print(dados_entradas_experiment.head())
```

	cod_sap	year_entrada	month_entrada	total_month_orders	max_variavel
37	110000018	2024	2	6	1
60	110000020	2024	1	7	1
61	110000020	2024	2	4	1
121	110000025	2023	2	4	1
122	110000025	2023	3	5	1

In [170]: `datos_entradas_experiment`

Out[170]:

	<code>cod_sap</code>	<code>year_entrada</code>	<code>month_entrada</code>	<code>total_month_orders</code>	<code>max_variavel</code>
<code>37</code>	<code>110000018</code>	<code>2024</code>	<code>2</code>	<code>6</code>	<code>1</code>
<code>60</code>	<code>110000020</code>	<code>2024</code>	<code>1</code>	<code>7</code>	<code>1</code>
<code>61</code>	<code>110000020</code>	<code>2024</code>	<code>2</code>	<code>4</code>	<code>1</code>
<code>121</code>	<code>110000025</code>	<code>2023</code>	<code>2</code>	<code>4</code>	<code>1</code>
<code>122</code>	<code>110000025</code>	<code>2023</code>	<code>3</code>	<code>5</code>	<code>1</code>
<code>...</code>	<code>...</code>	<code>...</code>	<code>...</code>	<code>...</code>	<code>...</code>
<code>12491</code>	<code>110088496</code>	<code>2023</code>	<code>12</code>	<code>5</code>	<code>1</code>
<code>12492</code>	<code>110088496</code>	<code>2024</code>	<code>1</code>	<code>5</code>	<code>1</code>
<code>12493</code>	<code>110088496</code>	<code>2024</code>	<code>2</code>	<code>7</code>	<code>1</code>
<code>12494</code>	<code>110088496</code>	<code>2024</code>	<code>3</code>	<code>6</code>	<code>1</code>
<code>12495</code>	<code>110088496</code>	<code>2024</code>	<code>4</code>	<code>5</code>	<code>1</code>

1058 rows x 5 columns