

**NATÁLIA CAETANO SENDRA
THIAGO DA SILVA COSTA**

**Declarative Process Mining
aplicado à evolução da COVID-19 na América Latina**

**PROJETO DE GRADUAÇÃO EM ENGENHARIA DE
PRODUÇÃO APRESENTADO AO DEPARTAMENTO DE
ENGENHARIA INDUSTRIAL DA PUC-RIO, COMO PARTE DOS
REQUISITOS PARA OBTENÇÃO
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO**

Orientadora: Fernanda Araujo Baião Amorim

**Departamento de Engenharia Industrial
Rio de Janeiro, 17 de novembro de 2023.**

Resumo

O presente estudo foi realizado com o intuito de analisar a pandemia da Covid-19 na América Latina através da mineração de processos declarativos. E, alguns bancos de dados disponíveis ao público foram utilizados: Google Mobility Data, banco mundial, organização internacional do trabalho, nações unidas, Oxford Covid Policy Tracker e Our World in Data. Nesse contexto, a pergunta de pesquisa é “Como as medidas farmacêuticas e não farmacêuticas impactaram a evolução da pandemia da COVID-19?”. Para responder essa questão, as ferramentas de mineração de processo foram utilizadas. Os dados coletados dos múltiplos bancos de dados foram adaptados para a análise ser viável. Consequentemente, os resultados principais obtidos foram a relação de suporte financeiro com a redução nas atividades nos locais de trabalho em momentos críticos, e a proteção de idosos, com a queda nas atividades nos parques em momentos de desaceleração. Por fim, foi comparado com a abordagem de mineração imperativa, as limitações foram pontuadas, assim como possíveis trabalhos futuros.

Palavras-chave: Mineração de Processos Declarativa, Covid-19, América Latina, Mineração de Processos Imperativa

Abstract

The present study was conducted to analyze the Covid-19 pandemic in Latin America through declarative process mining. Various publicly available databases were utilized, including Google Mobility Data, World Bank, International Labour Organization, United Nations, Oxford Covid Policy Tracker, and Our World in Data. In this context, the research question is "How did pharmaceutical and non-pharmaceutical measures impact the evolution of the COVID-19 pandemic?" To address this question, process mining tools were employed. The data collected from multiple databases were adapted to ensure the feasibility of the analysis. Consequently, the obtained results were the relationship of financial support with the reduction in activities in workplaces during critical moments and also the protection of the elderly, with the decline in activities in parks during slowdowns. Finally, when compared to the imperative approach, limitations were highlighted, as well as potential future work.

Keywords: Declarative Process Mining, Covid-19, Latin America, Imperative Process Mining

SUMÁRIO:

1-Introdução.....	1
2-Fundamentação Teórica.....	3
2.1-Ciência de dados.....	3
2.2-Mineração de Processos.....	6
2.3-Abordagens de Mineração de Processos e Ferramentas.....	8
2.3.1- Mineração de Processos Imperativa.....	8
2.3.1.1-Ferramentas	12
2.3.2-Mineração de Processos Declarativa.....	13
2.3.3-Análise comparativa Mineração de Processos Declarativa e Imperativa.....	19
3- Estudo de Caso.....	21
3.1- Cenário	21
3.2- Coleta de dados.....	26
3.3- Pré-processamento	28
3.4- Formatação Event log	37
3.5- Descoberta de cenários.....	38
3.5.1- Modelagem Declarativa	38
3.5.2- Modelagem Imperativa.....	45
4-Resultados.....	50
5-Conclusão.....	53
6-Referências.....	55

LISTA DE FIGURAS

Figura 1 –Ciclo de vida ciência de dados.....	5
Figura 2 –Exemplo de uma Petri Net.....	9
Figura 3 –Exemplo de um DFG.....	10
Figura 4 –Visão geral da abordagem de mineração de processos baseada em SQL.....	14
Figura 5 –Abordagens de Mineração de Processos.....	19
Figura 6 - Crescimento econômico na Região das Américas, 2010-2022 (%)	22
Figura 7 - Curva da média móvel de 7 dias de novos casos diários relativizados pelo tamanho da população dos 20 países da América Latina.....	24
Figura 8 - Mapa com os países analisados.....	25
Figura 9 - Teletrabalho e informalidade	27
Figura 10 - Variáveis de eventos e o significado de cada uma delas.....	35
Figura 11 - Etapas de extração e pré-processamento.....	36
Figura 12 - Log de eventos referente a eventos e medidas de contenção da COVID-19.....	38
Figura 13 - Diagrama Processo Celonis com data de início e ajustes.....	39

LISTA DE TABELAS

Tabela 1 – Dataset antes da formatação	29
Tabela 2 –Dataset depois da formatação	29
Tabela 3 - Dataset depois da concatenação das colunas “Eventos” e “Valor”	34
Tabela 4 - Principais regras do momento de aceleração	39
Tabela 5 –Principais regras do momento de desaceleração	40
Tabela 6 - Regras de alta confiança por país	42
Tabela 7 - Regras de baixa confiança por país	44
Tabela 8 - Duração média medida	45
Tabela 9 - Conexões do momento de aceleração, onde o processo segue do Evento 1 para o Evento 2	48
Tabela 10 - Conexões do momento de desaceleração, onde o processo segue do Evento 1 para o Evento 2.	48
Tabela 11 - Quadro comparativo das ferramentas	51

1 Introdução

A pandemia da COVID-19 afetou a população mundial na esfera da saúde, da economia e mais especificamente provocou um aumento da desigualdade. Além disso, a Organização Mundial da Saúde foi informada de mais de 500 milhões de casos e 6 milhões de mortes de Covid-19 até o ano de 2022. No entanto, esses são os dados oficiais e informados, então com a limitação das testagens e dos dados de saúde de alguns países os números podem ser bem maiores. (World Health Statistics 2022)

Além disso, o progresso tecnológico proporcionou ferramentas e recursos de grande poder computacional às empresas, permitindo a coleta e processamento de grandes volumes de dados em tempo recorde e escalas nunca vistas. Segundo Awotunde *et al.* (2022), a análise desses dados possibilita a obtenção de inúmeros insights valiosos em diversos contextos, inclusive para o cenário atual de pandemia da COVID-19; tais autores também enfatizam como as análises do grande volume de dados da epidemia de COVID-19 têm o potencial de auxiliar na tomada de decisões estratégicas e na contenção da disseminação do vírus, e salientam a possibilidade de, a partir dos dados, responder a inúmeras perguntas de forma prospectiva ou retrospectiva.

Do ponto de vista do tratamento, análise e extração de conhecimento a partir dos dados, e tendo em vista a enorme e crescente disponibilização de informações, a pandemia de COVID-19 deve ser considerada como possível referencial para o combate a futuras pandemias (MENEZES, 2022). Neste contexto, a motivação deste trabalho é analisar dados sobre os fatores que impactaram a quantidade de casos de COVID-19 de modo a entender o quanto seguir o processo padronizado era importante e se os países estavam certos em suas medidas.

Os acadêmicos, como Awotunde *et al.* (2022) e Jazbik (2022), se voltaram à elaboração de artigos sobre a pandemia pela sua importância e impactos. No entanto, a pandemia é recente e a produção de artigos com mineração de processos (process mining) relacionado a pandemia ainda está em uma fase mais embrionária. Já a área process mining e a área de saúde possuem muito mais conteúdo produzido e estabelecido. Porém, observou-se uma lacuna de conteúdo produzido voltado para a

América latina. . Então este artigo visa tentar preencher essa lacuna.

O presente estudo aplica técnicas de Process mining, uma área de conhecimento que engloba metodologias para descobrir, monitorar e melhorar processos reais de negócio. Esta área de conhecimento, como é conhecida atualmente, surgiu em 1990 e é a entre Process Science (Ciência do Processo) e Data Science (Ciência de Dados), e é empregada em muitos processos cotidianos (AALST, 2022a).

Assim como na abordagem tradicional (imperativa), as técnicas de Declarative Process Mining também englobam metodologias para a descoberta do processo, checagem da conformidade e monitoramento de conformidade (MAGGI, 2019). Além disso, os modelos declarativos são especialmente adequados para o âmbito da saúde devido à sua flexibilidade, o que os torna compatíveis com a imprevisibilidade e instabilidade inerentes aos processos do sistema de saúde (ROVAN *et al.*, 2015). O estudo de caso desenvolvido neste trabalho será feito utilizando algumas bases de dados cujo impacto na evolução da curva de casos da pandemia nos diferentes países analisados não está claro, dentre elas temos a Oxford Covid Policy Tracker, Google Mobility data e The World Bank. E, seguiu as etapas do ciclo de vida da ciência de dados. O objetivo final é ser capaz de responder a pergunta teórica: Como as medidas farmacêuticas e não farmacêuticas impactaram a evolução da pandemia da COVID-19?

Esperamos que este estudo contribua de forma positiva aos estudos da pandemia COVID-19 de forma a mitigar problemas em futuras pandemias ou crises na saúde e mostre a importância da análise e gestão dos processos para as atividades cotidianas.

Por fim, o estudo está dividido em 5 capítulos: o primeiro que é este de introdução, que apresenta a ideia do estudo e a pergunta a ser respondida, o segundo de fundamentação teórica que explora os conceitos importantes para o estudo de caso e para a escolha da metodologia, o terceiro do estudo de caso em si com suas análises e comparações, o quarto de resultados para apresentar o que foi obtido através do estudo e o quinto é a conclusão para levar ao entendimento final das ideias exploradas, as limitações e possíveis trabalhos futuros.

2 Fundamentação Teórica

O objetivo deste capítulo é apresentar os principais conceitos sobre mineração de processos, imperativa e declarativa e como elas se situam no domínio da saúde, além de auxiliar no entendimento dos conceitos necessários para o entendimento do estudo de caso.

2.1 Ciência de Dados

A ciência de dados é um campo interdisciplinar que tem por objetivo transformar dados em valores. Ela envolve diversas etapas desde a extração e preparação de dados até a apresentação de resultados, levando em consideração aspectos éticos, sociais, legais e de negócios (AALST, 2016).

Já Provost e Fawcett (2013) definiram que a ciência de dados consiste em um conjunto de princípios fundamentais que fornecem suporte e orientação para a extração cuidadosa de informações e conhecimento a partir de dados.

O ciclo de vida da pesquisa de ciência de dados pode ser descrito como contendo dois níveis, o externo e interno, onde o ciclo inteiro envolve 6 etapas cíclicas, descritas a seguir.

A primeira etapa é externa com o entendimento do problema, ela é a mais essencial, pois é sobre fazer as perguntas certas.

A coleta de dados inclui compreender os tipos de dados disponíveis. Isso abrange dados estruturados, exemplificados por tabelas de bancos de dados e arquivos CSV organizados em linhas e colunas; dados semi-estruturados, como logs e tweets sem uma estrutura fixa; e dados não estruturados, como fluxos de vídeo, imagens e sons. Além disso, existem duas categorias de fontes de dados: as internas, integradas aos sistemas corporativos e data warehouses, e as externas, como a NOAA (Administração Oceânica e Atmosférica Nacional) dos Estados Unidos, que fornecem dados adicionais para pesquisa, como previsões meteorológicas para gestão de energia.

A implementação de coletores de dados ou gerentes de dados é essencial nesse processo, garantindo não apenas a coleta, mas também a qualidade e a atualidade dos dados.

A terceira etapa inclui os procedimentos de análise de dados. O resultado desse ciclo na pesquisa em Ciência de Dados é um script ou software que executa uma sequência de comandos, desde o carregamento dos dados até a obtenção dos resultados desejados.

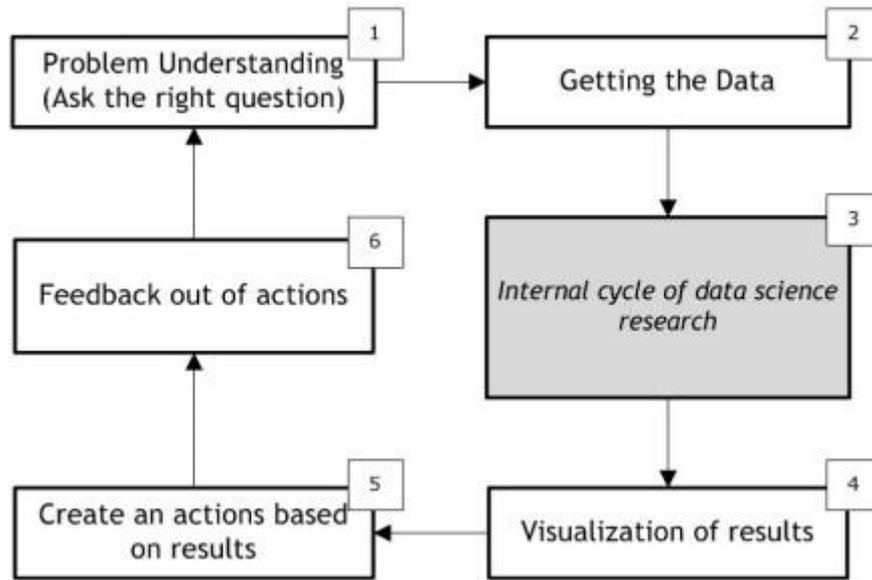
A quarta etapa do processo envolve a visualização dos resultados obtidos. Nesta fase, são exploradas diversas abordagens para representar os dados, abrangendo uma variedade de tipos de visualizações. Isso engloba a apresentação dos resultados como ponto de partida para apoiar a tomada de decisões, como a representação de séries temporais com valores previstos ao longo do horizonte de previsão ou gráficos de dispersão derivados da análise de agrupamento. Adicionalmente, são realizadas visualizações comparativas dos resultados com baselines ou modelos de referência, assim como a apresentação gráfica de informações cruciais à fase de modelagem, como AUC, ROC, erros de previsão e análise residual, entre outros.

A quinta é criar ações baseadas nos resultados, que pode ser positivo ou negativo, e em casos de resultados positivos, é necessário identificar e realizar ações com base nos resultados, como o sucesso da fase de pesquisa. Em contrapartida, em resultados negativos, é necessário repetir o ciclo de análise de dados e aprimorá-lo.

Por fim, a última etapa é obter feedback da ação. Esta etapa envolve a avaliação de como os resultados da análise foram úteis e confiáveis para o usuário final e a pessoa que toma decisões. A qualidade do suporte à decisão pode ser medida usando indicadores -chave de desempenho (KPIs).

Este ciclo de vida pode ser observado na figura 1 com as seis etapas de forma cíclica (SHCHERBAKOV *et al*, 2014).

Figura 1 – Ciclo de vida ciência de dados



Fonte: SHCHERBAKOV *et al*, (2014)

As abordagens tradicionais da literatura da ciência de dados focam no aspecto estrutural da realidade, onde o conhecimento é extraído por meio da descrição detalhada dos objetos, suas características e como eles podem se relacionar. Um exemplo disso é a utilização de informações de exames, sintomas, diagnóstico de doenças ou condições de saúde, conforme abordado por Wasan *et al*. (2006).

Por outro lado, uma abordagem complementar à ciência de dados que se tem feito na literatura é a Process Mining. Ela é uma abordagem bastante útil para se extrair conhecimento na perspectiva comportamental da realidade, ou seja, extraindo conhecimento a partir da evolução dos objetos ao longo de eventos ou processos que eles participem. Um exemplo disso é o estudo de caso desenvolvido em Rovani *et al*. (2015), onde se é examinado um processo de tratamento médico (composto por eventos do tipo intervenções médicas, administração de medicamentos, exames, etc...) para analisar as discrepâncias entre o comportamento esperado e real.

2.2 Mineração de Processos

Segundo Aalst (2022a), o process mining é a intersecção entre ciência de dados e ciência de processos, cujo objetivo é melhorar performances operacionais através do uso sistemático de dados de eventos. Já Erdogan e Tarhan (2018) são mais breves na sua definição, e descrevem process mining como uma técnica de gerenciamento de processos que explora dados de um sistema de informação. Ela pode ser aplicada principalmente em três situações:

- i) Descoberta de processo
- ii) Verificação de conformidade
- iii) Análise de desempenho

Descoberta de Processo é a atividade mais importante de process mining, dado que fornece a base para outras futuras análises envolvendo as outras situações de process mining.

Já na verificação de conformidade, o modelo do processo existente (considerado como sendo de referência) é comparado ao processo descoberto a partir dos registros de dados. Depois de realizada a comparação dos processos, uma análise de conformidade é feita, a fim de avaliar se o processo obtido a partir dos registros sistêmicos é compatível com o modelo do processo de referência.

Enquanto a análise de desempenho visa descobrir problemas (gargalos, atrasos) e oportunidades de melhoria do processo obtido a partir dos registros sistêmicos.

Tanto a descoberta de processo, como a verificação de conformidade e a análise de desempenho só podem ser feitos mediante ao input de dados. Esses dados são armazenados e processados pelos mais diversos sistemas de informação. Eles registram a informação detalhada sobre a identidade e a sequência de atividades que ocorreram na execução do processo. Essa informação é chamada de event log.

O log de eventos (event log) é geralmente formado pela:

i) instância: identificador da execução do processo. No domínio da saúde, pode ser, por exemplo, um paciente sendo tratado no hospital em um processo clínico. Em

um processo organizacional pode ser um pedido de venda, ao longo de todas as etapas do seu atendimento.

ii) evento: etapa ou atividade do processo. Então, no domínio da saúde, um exemplo de evento pode ser 'Verificar sinais vitais' em um processo clínico ou 'Enviar lembrete de pagamento' em um processo organizacional.

iii) timestamp: data do registro do evento.

Então, uma vez tendo o event log de um determinado processo é possível aplicar as técnicas de process mining.

No entanto, como pontuam Munoz-Gama *et al.* (2022), no domínio da saúde, os dados de registros costumam sofrer com problemas de qualidade. Em Rojas *et al.* (2016), se acrescenta outro aspecto negativo com relação ao uso do process mining: a falta de uma boa visualização dos modelos de processo e do resultado obtido, o que acaba prejudicando bastante o entendimento do que foi gerado pelas ferramentas, ainda mais se tratando do domínio da saúde que possui processos complexos, dinâmicos e poucos estruturados.

Mertens *et al.* (2017) propôs a aplicação de declarative process mining para auxiliar na representação de processos flexíveis, dinâmicos e complexos. Segundo Rovani *et al.* (2015) declarative process mining é uma abordagem para extrair modelos de processos de eventos registrados em um log, onde as restrições são expressas usando uma linguagem declarativa. Mertens *et al.* (2017) propõe uma técnica que permite que os usuários especifiquem suas intenções de análise em termos de objetivos e restrições, ao invés de explicitamente modelar o processo de saúde. Essa abordagem reduz a complexidade do processo de modelagem e permite que os usuários analisem os dados de saúde de uma forma mais intuitiva.

Seja o imperative process mining ou o declarative process mining, a mineração de processos como um todo é fundamental especialmente na identificação de gargalos e oportunidades para otimização de processos de saúde. A pandemia de COVID-19 trouxe desafios significativos para a área da saúde, incluindo a necessidade de uma resposta rápida e eficaz para controlar a disseminação do vírus e garantir a disponibilidade de recursos médicos. Nesse contexto, o process mining pode ser uma ferramenta valiosa para identificar gargalos e otimizar a alocação de recursos em tempo

real.

2.3 Abordagens de Mineração de Processos e Ferramentas

Process Mining pode ser aplicado segundo diversas abordagens. Neste subtópico serão analisadas as abordagens imperativa e declarativa, e suas respectivas ferramentas, além de demonstrar como a declarativa é a mais adequada para a área da saúde através de uma análise comparativa.

2.3.1 Mineração de Processos Imperativa

A abordagem imperativa de mineração de processos é a que produz representações fechadas, ou seja, apenas as maneiras de execução previstas no modelo são válidas. Os modelos imperativos são adequados para descrever explicitamente padrões de fluxo de controle, como sequências, escolhas, seções concorrentes e loops. Esses padrões são eficazes na representação de uma parte dos caminhos possíveis, mas encontram dificuldades ao lidar com a gama completa de caminhos de execução em processos pouco estruturados e flexíveis. Em resumo, esses modelos dão ênfase ao controle em detrimento da flexibilidade (DI CICCIO e MONTALI, 2022).

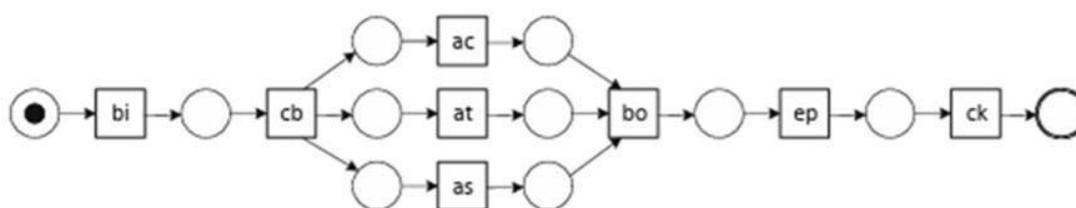
Os processos podem ser representados na forma imperativa através de diversas notações diferentes, sendo as quatro principais: BPMN (Business Process Model and Notation), Petri net, DFG (Directly Follow Graphs) e Process Tree. Cada processo requer uma representação específica que seja mais adequada às suas características. Por exemplo, processos altamente variáveis podem não ser adequadamente representados por meio do BPMN.

Primeiramente, o BPMN, também conhecido por Business Process Model and Notation é a representação padrão para BPM (Business Process Modeling) na indústria. A notação é relativamente complexa com seus mais de 100 símbolos. Essa notação engloba eventos, atividades e arcos, onde eventos são representados por círculos, atividades por retângulos e os arcos por setas (DUMAS *et al.*, 2018). O ponto fraco está na complexidade, que pode ser restritiva para alguns usuários e inadequada para

processos mais simples.

Petri Net é uma notação bastante visual que auxilia na observação do estado do modelo do processo. Isso ocorre pois, antes de cada transição que modela a atividade, existe um círculo que modela o estado, que quando contém um token (representado por um ponto preto), está habilitada a transição para o próximo estado (DI CICCIO e MONTALI, 2022).

Figura 2 - Exemplo de uma Petri Net



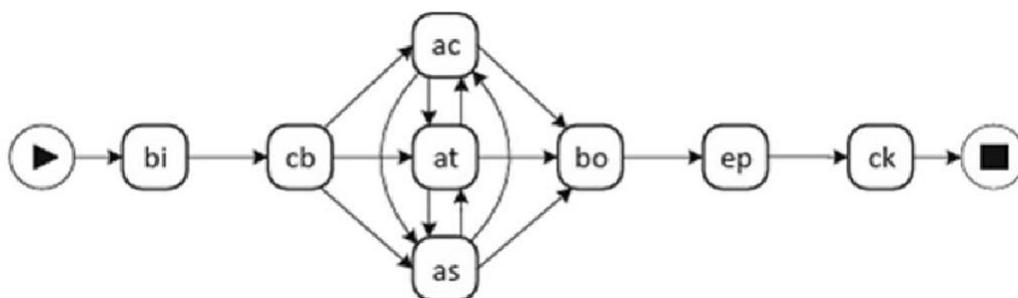
Fonte: AALST, (2022a)

Uma rede de Petri é um modelo lógico que não tem a intenção de representar o tempo de ocorrência de eventos, mas sim a ordem em que os eventos ocorrem. As Petri Nets foram desenvolvidas especialmente para modelar sistemas com componentes que interagem, e, portanto, são capazes de abranger diversas características de um sistema orientado por eventos, tais como concorrência, operações assíncronas, impasses, conflitos, entre outros (CABASINO *et al.*, 2013).

DFG (Directly Follow Graphs) é uma notação que mostra a sequência possível imediata do passo anterior através do arco entre os passos, conhecido por directly-follows relation. Desse modo, o DFG consiste em nós de início, fim e atividades estando todos atrelados aos arcos directly-follows relation. Então, a maior parte das ferramentas de Process Mining mostram diretamente esta notação ao carregar um registro de eventos por ser a mais rápida e simples de visualizar. Segundo (AALST, 2022), o uso de DFGs geralmente gera modelos de processos com subajuste. E, outro dos problemas dessa notação consiste na ocorrência frequente de loops, pois é impossível modelar corretamente a simultaneidade das atividades com essa notação que não possui uma

ordem estrita. E, com um loop desses na modelagem, existem infinitas maneiras de realizar o processo (AALST, 2022a).

Figura 3 -Exemplo de um DFG



Fonte: AALST, (2022a)

Por último, Process Tree é uma notação raramente usada com o usuário final, mas sim internamente em muitos algoritmos de mineração. O seu nome vem do fato de ser parecida com uma árvore e sua notação é uma seta indicando a direção da ocorrência e vários ramos com as atividades a serem representadas. Então, é sequencial, e em ramos com mais de uma atividade elas são paralelas ou exclusivas (DI CICCIO e MONTALI, 2022).

Além das diferentes notações que podem ser geradas para representar um modelo de processo, a mineração em si pode ser realizada segundo variados algoritmos como Alpha Miner, Heuristic Miner e Inductive Miner. Os algoritmos Alpha Miner de 2002 e Heuristic Miner de 2003 são ambos de descoberta de processos (GARCIA *et al.*, 2019).

O Alpha Miner é utilizado para reconstruir a causalidade a partir de sequências de eventos e é definido em termos de uma Petri Net. Ele investiga a perspectiva do fluxo de controle de um processo, identificando as relações de dependência entre suas tarefas. No entanto, este algoritmo assume que o registro minerado deve ser completo e livre de ruídos, o que nem sempre é viável na prática. Além disso, o Alpha Miner não utiliza informações de frequência (POROUHAN *et al.*, 2014).

Já o Heuristic Miner (HM) apresenta três melhorias significativas em relação ao algoritmo Alpha. Primeiramente, considera frequências e significância, o que permite

filtrar comportamentos pouco frequentes e ruidosos, tornando-o menos sensível a registros incompletos. Em segundo lugar, é capaz de detectar loops curtos, e, por último, permite a omissão de atividades individuais (BOGARÍN *et al.*, 2018). O método clássico do Heuristic Miner envolve cinco etapas principais para criar um modelo de processo que representa o comportamento de um registro de eventos, incluindo a identificação da matriz de pegadas, o cálculo das medidas de dependência, a elaboração de um gráfico, a descoberta de divisões e junções (representados pelos gateways lógicos), e o ajuste da mineração de loops de comprimento 1 e 2 (ARAGHI *et al.*, 2022).

O Inductive Miner é um algoritmo valorizado na descoberta de modelos de processos, identificando gateways frequentes em registros de eventos e aplicando limiares (thresholds) para atividades e conexões, resultando em diversos modelos. Um exemplo disso é o plugin Inductive Visual Miner no ProM3 (ARAGHI *et al.*, 2022). Embora o Inductive Miner produza modelos sólidos, ele enfrenta desafios na representação de comportamentos complexos, muitas vezes gerando modelos imprecisos, conhecidos como "modelos florais". Além disso, sua capacidade de lidar com comportamentos de alternância é limitada. O Inductive Miner segue uma abordagem de divisão e conquista, culminando em árvores de processos. Embora garanta solidez, não consegue representar todos os tipos de comportamentos. Algumas tentativas, como a duplicação de rótulos, foram feitas para aprimorar a precisão, mas isso resulta em maior tempo de execução. O framework original do Inductive Miner oferece a flexibilidade de personalizar a definição de atividades e a semântica das árvores de processos, permitindo a distinção entre comportamentos de "entrelaçamento" e "paralelismo" (LU *et al.*, 2021).

2.3.1.1 Ferramentas

A mineração de processos imperativa também possui diversas ferramentas que se utilizam dessas notações e algoritmos como a ProM, Disco, Celonis e PowerAutomate (GOMES *et al.*, 2021).

O ProM é um framework de código aberto que suporta diversos algoritmos de mineração de processos por meio de plugins. Ele é amplamente utilizado e flexível em termos de formatos de entrada e saída, incluindo redes de Petri e redes sociais. Os plug-

ins se dividem em cinco categorias: mineração, exportação, importação, análise e conversão. Eles oferecem funcionalidades variadas, como visualização, armazenamento, processamento de dados e análise, tornando o ProM uma ferramenta versátil para análise de processos. Além disso, sua grande vantagem é na verificação de conformidade por essa parte ser bastante intuitiva e efetiva na ferramenta (GOMES *et al.*, 2021).

O Disco é uma solução proprietária com funcionalidades avançadas para mineração de processos. Ele foi projetado para tornar a importação de dados simples e rápida, permitindo a detecção automática de campos em registros. Após a importação, o Disco apresenta automaticamente um modelo de processo que mostra visualmente como as atividades foram executadas. Ele fornece informações estatísticas, como frequência e desempenho das atividades, além de permitir a análise de variantes específicas do processo. O Disco oferece recursos de filtragem para explorar diferentes aspectos do processo e responder a perguntas específicas. Existem cinco tipos de filtros disponíveis: de intervalo de tempo, de variação, de desempenho, de pontos de extremidade e de atributos (GOMES *et al.*, 2021).

Celonis, por sua vez, é uma ferramenta de análise de processos de negócios que utiliza mineração de processos e inteligência artificial para extrair insights valiosos dos dados operacionais de uma organização. Possibilita a visualização detalhada dos processos, identificação de gargalos e oportunidades de melhoria. A plataforma importa dados de várias fontes, como logs de sistemas e registros de transações, processando-os em tempo real para oferecer uma visão imediata dos processos. Utilizando algoritmos avançados, identifica padrões e anomalias, revelando insights ocultos. Com base nessas informações, o Celonis capacita a automação de tarefas, otimização de fluxos de trabalho e implementação de melhorias contínuas. Oferece recursos de monitoramento em tempo real, alertas e painéis interativos para acompanhar o desempenho dos processos, proporcionando às organizações a oportunidade de impulsionar a eficiência operacional, reduzir custos, aumentar a produtividade e melhorar a experiência do cliente (VOGELGESANG *et al.*, 2022).

No entanto, os mapas gerados pelo Celonis, Disco e alguns plugins do ProM podem ser considerados difíceis de interpretar. Em alguns casos, essas ferramentas podem até falhar em produzir mapas legíveis. Recomenda-se ajustar os parâmetros das ferramentas, especialmente no Celonis e no Perceptive Process Mining, para melhorar

a legibilidade dos mapas. As ferramentas comerciais Celonis, Disco e ProM oferecem interfaces de usuário amigáveis, com recursos como feedback de parâmetros, animação de logs e opções de filtragem. No entanto, observou-se que os mapas de processos criados por essas ferramentas muitas vezes não apresentam paralelismo e semântica executável, resultando em cálculos imprecisos de desvios (KESICI *et al.*, 2022).

Por fim, a abordagem da Microsoft para a mineração de processos utiliza dados de eventos extraídos dos sistemas de registro para criar um modelo dos processos e mostrar como os dados fluem entre esses sistemas. O Power Automate Process Mining exige acesso aos logs do aplicativo, podendo ser uma conexão direta ou exportação em CSV. Após obter acesso aos dados de log no Power Automate, a ferramenta utiliza o Editor do Power Query para transformar dados, identificar atividades e rastrear os processos de negócios. Uma aplicação desktop oferece recursos de edição de processos, gerando mapas que indicam a descoberta, execução e repetição de atividades. Os cronogramas mostram a duração total e de cada atividade, auxiliando na priorização e otimização de processos. Indicadores-chave de desempenho resumem dados para identificar oportunidades de melhoria rápida, como a redução de loops em um processo (BISSON, 2023).

2.3.2 Mineração de Processos Declarativa

Nesse item, será apresentada a abordagem declarativa, que será a parte principal do estudo de caso e é muito relevante no contexto da área da saúde.

Declarative Process Mining (Mineração de Processos Declarativa) é a abordagem de mineração de processos na qual o processo é representado como conjunto de restrições, logo respeitando as restrições o processo pode ser executado de todas as maneiras possíveis (MAGGI *et al.*, 2011). Dessa forma, a abordagem declarativa define de forma implícita o fluxo de controle como todas as possibilidades que não violam nenhuma das restrições (PESIC *et al.*, 2007).

Já DI CICCIO e MONTALI (2022) aborda Declarative Process Mining como capaz de capturar o que é esperado que ocorra, mas também captura o que não deve ocorrer. As restrições de processos declarativos refletem como cada atividade interage

com as demais. Essa análise permite examinar os relacionamentos comportamentais de cada atividade no contexto do processo como um todo. Em outras palavras, as restrições de uma atividade individual influenciam o comportamento global do processo.

Existem três aplicações principais na mineração de processos declarativa, as mesmas de mineração de processos: descoberta de processo, verificação de conformidade e análise de desempenho. Uma abordagem declarativa permite que os analistas de negócios se concentrem em algumas restrições importantes e estáveis ao longo do tempo, evitando a complexidade de modelar muitos detalhes de fluxo de controle suscetíveis a mudanças. Isso mantém os processos mais estáveis, permitindo que algumas regras acomodem várias trajetórias de execução. Essa abordagem é particularmente útil em processos de saúde, que são altamente variáveis (MAGGI, 2019).

Declare, por sua vez, é uma linguagem e notação gráfica que oferece uma ampla variedade de modelos extensíveis para a formulação de restrições e é a mais utilizada de Declarative Process Mining. Esta linguagem se origina a partir de 3 linhas de pesquisa, cada uma delas focadas na especificação declarativa de processos de negócios, coreografia de serviços e diretrizes clínicas. Os modelos essenciais do Declare são desenvolvidos com base em um catálogo de templates de lógica temporal utilizados na verificação de modelos em diversos sistemas dinâmicos de diferentes domínios de aplicação. (DI CICCIO e MONTALI, 2022). Os templates Declare utilizados no presente trabalho serão apresentados mais adiante, na seção do estudo de caso.

A abordagem declarativa na mineração de processos começou a se desenvolver em 2011 com a introdução do Declarative Miner, e ao longo do tempo, essa abordagem foi aprimorada e mais profundamente explorada (MAGGI *et al.*, 2011). Alguns dos algoritmos abordados nesta área incluem o Declarative Miner e MINERful.

O Declarative Miner foi o pioneiro na categoria de mineradores declarativos. Em sua versão inicial, ele utilizou uma estratégia de força bruta para identificar restrições declarativas. Isso envolveu a geração de todas as possíveis restrições com base nas atividades observadas no registro, a criação de fórmulas LTL, lógica temporal linear, para essas restrições e, em seguida, a verificação de quantas das trilhas no registro satisfaziam cada uma dessas restrições.

No ano seguinte, essa abordagem foi aprimorada com diversas melhorias

significativas. Uma das melhorias mais notáveis foi o aumento da eficiência do algoritmo, obtido ao considerar apenas as restrições que tinham maior probabilidade de serem interessantes. Isso foi possível graças a um processo de pré-processamento inspirado no algoritmo Apriori para mineração de regras de associação. Esse refinamento permitiu que o Declarative Miner fosse executado de maneira eficaz, mesmo em registros extensos com um grande número de atividades únicas, o que implicava em um vasto conjunto de possíveis restrições. Mais recentemente, em uma versão posterior, o minerador foi expandido para suportar a execução em paralelo (SLAATS, 2020).

Em 2013, foi introduzido o algoritmo MINERful, com foco principal em melhorar a eficiência. Ele adota uma abordagem heurística na descoberta de modelos Declare, construindo um banco de dados com abstrações do registro. Essas abstrações são usadas diretamente para identificar quais restrições Declare são cumpridas, tornando o algoritmo significativamente mais rápido do que o Declare Miner, que verifica todas as trilhas para cada restrição.

Uma comparação entre os modelos gerados pelo MINERful e o Declare Miner mostrou que o último tende a produzir modelos de processo mais adequados e concisos. E, desde sua criação, o MINERful foi aprimorado com suporte para restrições ramificadas, eliminação de redundâncias e inconsistências, bem como a remoção de restrições que eram cumpridas de forma vazia (SLAATS, 2020).

A área de Declarative Process Mining oferece várias ferramentas já desenvolvidas, disponíveis em Python, Java e SQL. Uma dessas ferramentas é o Declare4py, uma biblioteca Python voltada para o desenvolvimento de abordagens declarativas. Esta biblioteca foi criada com o propósito de analisar logs de eventos utilizando modelos de processos baseados em restrições, e para preencher a lacuna da falta de ferramentas em Python para a abordagem declarativa. Além disso, o Declare4py oferece três funcionalidades principais: conformance checking, process discovery e query checking. Por ser uma ferramenta mais recente ainda tem atualizações propostas para desenvolver e novas funcionalidades a implementar (DONADELLO *et al.*, 2022).

Na ferramenta de mineração de processos ProM, estão disponíveis diversos plugins que podem ser empregados para distintos tipos de análises baseadas em Declare, abrangendo desde a descoberta de modelos Declare até a verificação de

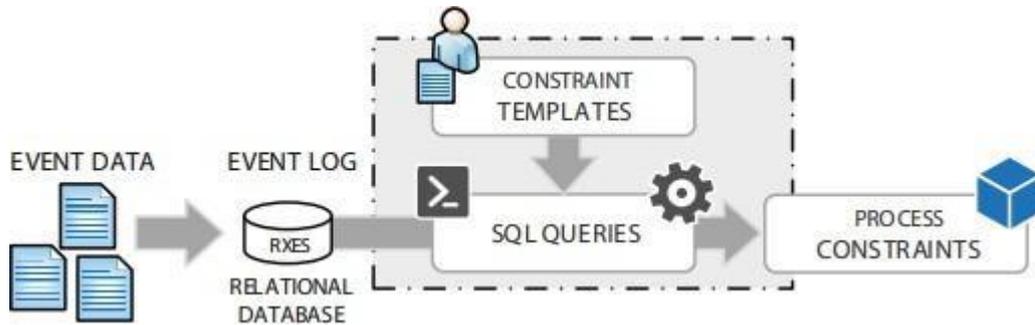
conformidade e o monitoramento online de instâncias de processos em execução. Existem três plugins principais voltados exclusivamente para a abordagem declarativa, eles são Declare Analyzer, Declare Checker e Declare Maps Miner. O Declare Checker pode ser empregado para verificar a conformidade do comportamento efetivo de um processo, conforme registrado em um log de eventos, em relação a um modelo Declare de referência. Por meio do Declare Maps Miner, é possível criar um conjunto de restrições Declare do zero, as quais representam o comportamento real de um processo conforme registrado em um registro de eventos. Por último, através Declare Analyzer, o usuário tem a capacidade de identificar precisamente onde a execução do processo se diferencia do modelo Declare de referência e quantificar o grau de conformidade do comportamento do processo utilizando diversas métricas, como a taxa de cumprimento e a taxa de violação (MAGGI, 2013).

Por fim, SQL queries também são utilizadas para extrair as regras declarativas. O estudo de caso vai se utilizar de diversas dessas queries para tentar estabelecer o relacionamento das variáveis através das regras e observar assim o comportamento das variáveis.¹

A técnica de mineração de processos através de consultas SQL, desenvolvida por Schonig *et al.*, 2016, oferece uma abordagem abrangente e integrada para a descoberta de processos. Ao aproveitar a tecnologia de desempenho de banco de dados relacional, ela é rápida e flexível, não se limitando a restrições predefinidas. As consultas podem ser adaptadas para vários aspectos de um processo, como fluxo de controle e questões organizacionais, e os resultados podem ser transformados para diferentes linguagens de modelagem de processos. A avaliação de desempenho foi realizada utilizando logs de eventos reais, e a abordagem demonstrou capacidade, expressividade e insights adicionais ao fornecer consultas para descobrir restrições de processo comuns. Em resumo, essa técnica é personalizável, independente de linguagem e eficiente, apresentando uma abordagem declarativa para a mineração de processos.

¹ Outras ferramentas foram testadas para o estudo de caso, porém apenas as queries SQL da abordagem declarativa foram utilizadas. A Declare4Py foi testada, porém o código apresentou falhas e por isso não foi utilizada.

Figura 4- Visão geral da abordagem de mineração de processos baseada em SQL



Fonte: SCHONIG *et al*, 2016

Essa figura ilustra esse processo de mineração de SQL queries onde se tem event data(dados de evento), que são agrupados em uma relational database (base de dados relacional) e são registrados como eventos (event log) onde as SQL Queries vão agir através dos regras de restrições(constraint regras) e por fim resultar nas restrições do processo.

As principais restrições da linguagem Declare são:

- I. Response: Esta restrição representa que "se a atividade A ocorre, então em algum momento do processo (não necessariamente logo em seguida) a atividade B ocorre após A".
- II. Alternate Response: Esta restrição representa que "se a atividade A ocorre, então em algum momento do processo (não necessariamente logo em seguida) a atividade B ocorre após A, sem outra ocorrência de A entre A e B".
- III. Alternate Precedence: Esta restrição representa que "se a atividade B ocorre, então a atividade A ocorre antes de B".
- IV. Precedence: Esta restrição representa que "se a atividade B ocorre, então a atividade A ocorre antes de B, sem outra ocorrência de B entre A e B".
- V. Responded Existence: Esta restrição representa que "se a atividade A ocorre, então a atividade B ocorre, antes ou depois de A".
- VI. Not Succession: Esta restrição representa que "a atividade A e a atividade B nunca ocorrem em sucessão".

Tais restrições foram trabalhadas no log de eventos do presente estudo de caso utilizando o MS SQL e as queries desenvolvidas por Schonig *et al.*, (2016).

As métricas utilizadas foram o Support e o confidence. Em Schonig *et al.*, (2016) essas métricas são definidas da seguinte maneira:

- I. Support: O suporte é a medida de quantas vezes uma determinada restrição (constraint) é satisfeita em relação ao número total de ocorrências da condição da restrição. Exemplo: Considere a restrição Response(a,b). Se esta restrição é satisfeita (ou seja, se “a” então “b” em resposta de “a”), o suporte seria o número de vezes que isso ocorreu dividido pelo número total de ocorrências de “a”.
- II. Confidence: A confiança é o produto do suporte e da fração de Instâncias no log onde a condição (implicações) ou a atividade restrita (não implicações) ocorre. Exemplo: Para Response(a,b), a confiança seria o suporte multiplicado pela fração de Instâncias em que a condição “a” ocorre. Isso reflete a probabilidade de que, dado ocorrência de “a”, a restrição será satisfeita.

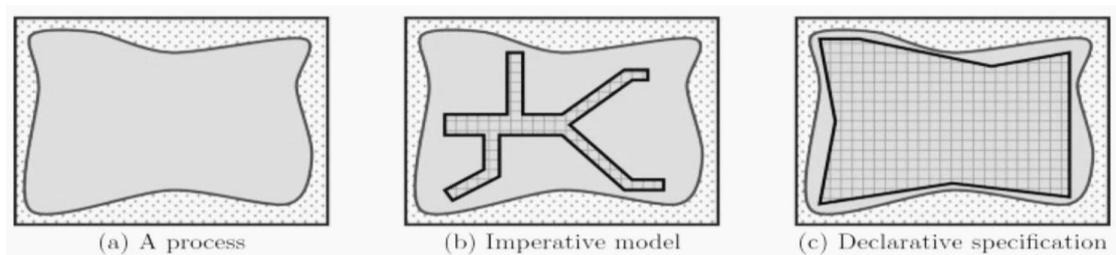
Em suma, o suporte é uma medida que avalia a frequência com que uma dada restrição é satisfeita, calculando a proporção entre o número de vezes que duas atividades ocorrem juntas (TaskA e TaskB, por exemplo) e o número total de vezes que TaskA ocorre. A confiança fornece uma indicação de quão forte é a associação entre as duas tarefas, dado que TaskA ocorreu.

2.3.3 Análise comparativa Mineração de Processos Declarativa e Imperativa

Em Back *et al.*, (2020) as notações de modelagem de processos podem ser divididas em duas amplas categorias: Declarativa ou Imperativa. Os conceitos de ambas foram expostos ao longo deste capítulo e neste subtópico iremos explorar e comparar suas diferenças.

Uma representação gráfica que ilustra as diferenças entre as abordagens é:

Figura 5 – Abordagens de Mineração de Processos



Fonte: DI CICCIO e MONTALI, (2022)

A figura mostra três diagramas. O diagrama (a) A process representa um processo no que limita no espaço cinza o comportamento permitido e fora na parte pontilhada o proibido. O seguinte (b) Imperative model representa o modelo imperativo que além da limitação já estabelecida ainda é bem mais reduzido na quantidade de comportamentos permitidos, então apesar de serem permitidos ficam de fora e só as execuções de processos previstas no modelo serão permitidas. Em último, o diagrama (c) Declarative specification, na direita, representa o modelo declarativo que fica dentro, mas bem próximo, dos limites dos comportamentos permitidos e engloba todo o modelo imperativo.

Então, a abordagem imperativa favorece controle em detrimento da flexibilidade, ao mesmo tempo a declarativa se esforça para balancear flexibilidade e controle tentando exprimir restrições que separam bem os comportamentos permitidos dos proibidos. (DI CICCIO e MONTALI, 2022)

Outro ponto relevante a destacar nesta comparação são os avanços recentes e futuras perspectivas e modelos de paradigma misto da abordagem declarativa (DI CICCIO e MONTALI, 2022). Ao longo da explicação do modelo é argumentado que a diferença entre as abordagens nem sempre é tão nítida e que um mesmo processo pode possuir partes mais bem descritas através do Imperative e outras através do Declarative. O exemplo abordado é da área da saúde onde se tem uma diretriz clínica misturando subprocessos administrativos com terapêuticos. Então, para representar processos híbridos é necessária uma abordagem multiparadigma capaz de integrar modelos imperativos e declarativos em um único modelo de processo. Por último, conclui que no estado atual é necessária mais pesquisa para desenvolver esse modelo.

Por fim, no próximo capítulo será proposto e aplicado o estudo de caso e visto

por ambas as Abordagens.

3 Estudo de Caso

3.1 Cenário

Como relata Pestel (2020), em dezembro de 2019 a Organização Mundial da Saúde (OMS) foi alertada sobre a ocorrência de inúmeros casos de pneumonia na cidade de Wuhan, na China. Os casos tinham um padrão incomum, o que levou as autoridades a investigarem as causas. Em janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de coronavírus. Esse novo coronavírus (nomeado como SARS-CoV-2 em fevereiro de 2020) é o vírus responsável por causar a doença COVID-19.

Ainda em janeiro de 2020, a OMS declarou o surto do novo coronavírus como uma Emergência de Saúde Pública de Importância Internacional (ESPII), o mais alto nível de alerta da Organização, como posto por Pestel (2020). Com essa decisão, a OMS iniciou esforços na direção de um aprimoramento da coordenação, cooperação e solidariedade global com finalidade de interromper a propagação do vírus.

Conforme diz Annelies e Osman (2020), a COVID-19 marca a sexta vez, no período entre 2007 e 2020, que uma Emergência de Saúde Pública de Importância Internacional é declarada. As outras ocorrências foram:

1. 25 de abril de 2009: pandemia de H1N1
2. 5 de maio de 2014: disseminação internacional de poliovírus
3. 8 agosto de 2014: surto de Ebola na África Ocidental
4. 1 de fevereiro de 2016: vírus zika e aumento de casos de microcefalia e outras malformações congênitas
5. 18 maio de 2018: surto de ebola na República Democrática do Congo

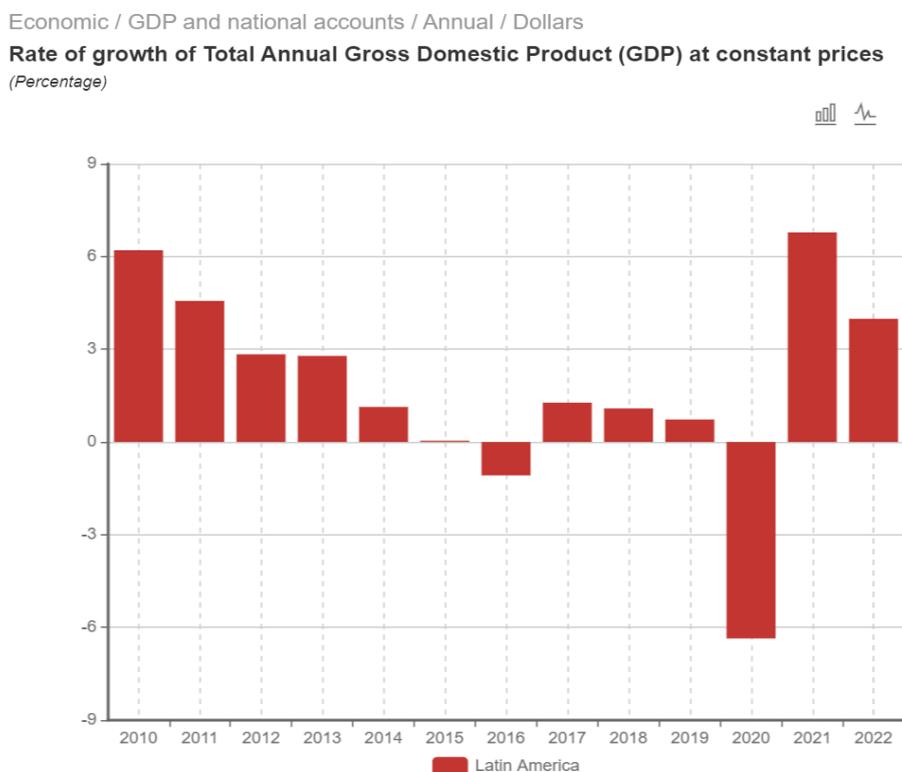
Em março de 2020, a COVID-19 foi caracterizada pela OMS como uma pandemia, termo que se refere à distribuição geográfica de uma doença e não à sua gravidade. A designação reconhece que, no momento, existem surtos de COVID-19 em vários países e regiões do mundo, conforme explica a Organização Pan Americana de

Saúde.

Nesse cenário, segundo a Organização Pan Americana de Saúde, em decorrência da pandemia, a região das Américas sofreu impactos: preços elevados das matérias-primas e as interrupções nas cadeias de suprimentos internacionais, o que culminou em um aumento dos custos. Tais fatores afetaram a economia da região, uma vez que contribuíram para a contração do crescimento econômico, índices elevados de inflação e um aumento nas taxas de desemprego, pobreza e insegurança alimentar.

Ainda de acordo com a Organização Pan Americana de Saúde, no âmbito do crescimento econômico, a pandemia e as medidas para contê-la causaram a pior recessão econômica global desde meados do século XX. A América Latina foi uma das áreas mais afetadas em termos socioeconômicos, experimentando uma queda de 6,8% no PIB em 2020, uma queda que superou as de outras partes do mundo. No entanto, em 2021, houve uma recuperação significativa com um crescimento de 6,8% no PIB devido à melhoria da situação global e à vacinação contra a COVID-19. Em 2022, o crescimento foi mais contido, declinando para 3,6%, como mostra a figura 6:

Figura 6 - Crescimento econômico na Região das Américas, 2010-2022 (%)



Source: CEPALSTAT - ECLAC - UNITED NATIONS

Fonte: Comissão Econômica para a América Latina e o Caribe. CEPALSTAT.

Principais Figuras da América Latina e do Caribe. Santiago: CEPAL; 2022.
Disponível em: <https://statistics.cepal.org/portal/cepalstat/index.html?lang=en>

Como mostra a Organização Pan Americana de Saúde, no que se refere às taxas de inflação, no segundo semestre de 2021, a América Latina e o Caribe enfrentaram uma taxa de inflação de 6,6%, o que foi um nível sem precedentes desde outubro de 2008. Essa alta foi observada em toda a região, com exceção de Argentina, Bolívia, Haiti, Suriname e Venezuela, que já tinham inflação elevada antes da pandemia. O aumento das taxas de juros nos países mais desenvolvidos e as complexas condições econômicas globais aumentam os riscos financeiros e as vulnerabilidades, especialmente para as economias da América Latina e do Caribe.

A alta inflação, ainda segundo a Organização Pan Americana de Saúde, gera incerteza financeira e pode reduzir o foco em investimentos em programas sociais e de saúde, pois os países tendem a priorizar o controle da inflação. No entanto, medidas para conter a inflação podem causar insatisfação social. Portanto, é importante que tais medidas incluam apoio adicional para proteger as populações vulneráveis.

De acordo com a Organização Pan Americana de Saúde, no quesito pobreza, pobreza extrema e distribuição de renda, antes mesmo da pandemia começar, a região das Américas enfrentava problemas. A pandemia exacerbou as desigualdades existentes na Região, aumentando o fosso. Na América Latina e no Caribe, a pobreza estava em ascensão, passando de 27,8% em 2014 para 33,7% em 2022, resultando em mais 22 milhões de pessoas vivendo abaixo da linha da pobreza, com 8 milhões (36%) em extrema pobreza. A COVID-19 acentuou essa tendência em 2020. Durante 2021, devido à recuperação econômica na Região, pode-se perceber uma redução da taxa de pobreza à medida que a cobertura vacinal contra a COVID-19 aumentava.

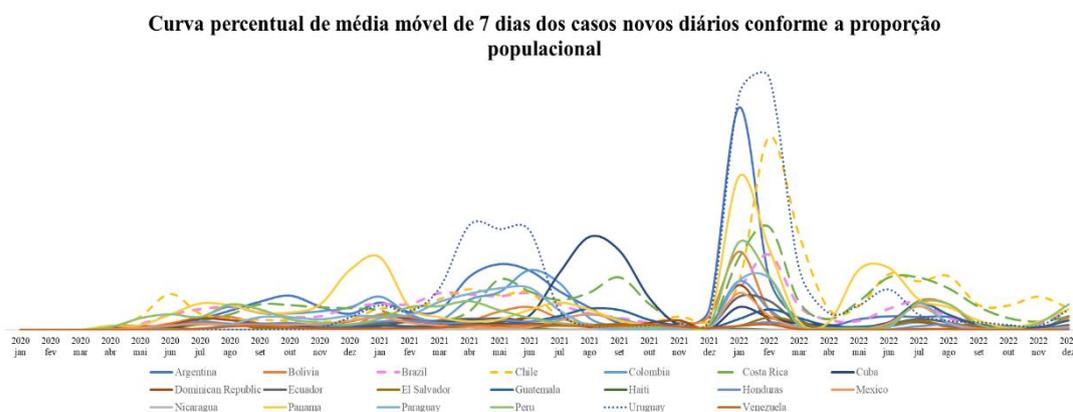
Como relata a Organização Pan Americana de Saúde, durante a pandemia, o fechamento das escolas acarretou a interrupção dos programas de alimentação escolar, levando ao aumento da insegurança alimentar que, por sua vez, pode causar impactos a longo prazo na saúde, produtividade, desenvolvimento, comportamento e aprendizado de crianças e adolescentes na região. As escolas desempenham um papel fundamental que vai além da educação, pois também protegem os direitos das crianças e contribuem

para o bem-estar social e emocional, que foi afetado pelo período de confinamento em casa.

Diante desse cenário, conforme explica a Organização Pan Americana de Saúde, inicialmente, os países da América Latina adotaram algumas respostas bastante previsíveis frente à pandemia. Declarações nacionais de emergência foram emitidas, grupos técnicos multissetoriais foram estabelecidos, e as agências de saúde foram reforçadas para enfrentar a crise. A maioria dos governos declarou a suspensão de atividades não essenciais, o fechamento de fronteiras e estabelecimentos como escolas, universidades e locais de trabalho. Dependendo do país, medidas adicionais foram tomadas, como o uso obrigatório ou voluntário de máscaras em locais públicos e fechados.

Embora exista certa similaridade nos países da América Latina, a pandemia afetou cada país de maneira muito particular. A figura 7 mostra a curva da média móvel de 7 dias de novos casos diários relativizados pelo tamanho da população dos 20 países da América Latina, onde fica evidente essa diferença entre os países.

Figura 7 - Curva da média móvel de 7 dias de novos casos diários relativizados pelo tamanho da população dos 20 países da América Latina.



Fonte: Elaboração própria

Pode-se perceber que, em dados momentos, determinados países alcançaram picos elevados, enquanto outros não. Mas vale notar que no final de 2021 muitos países tiveram valores altos. Esse momento coincide com o surgimento de uma variante da SARS-CoV-2, a Ômicron.

Como é possível notar no gráfico acima, a curva da média móvel de 7 dias dos

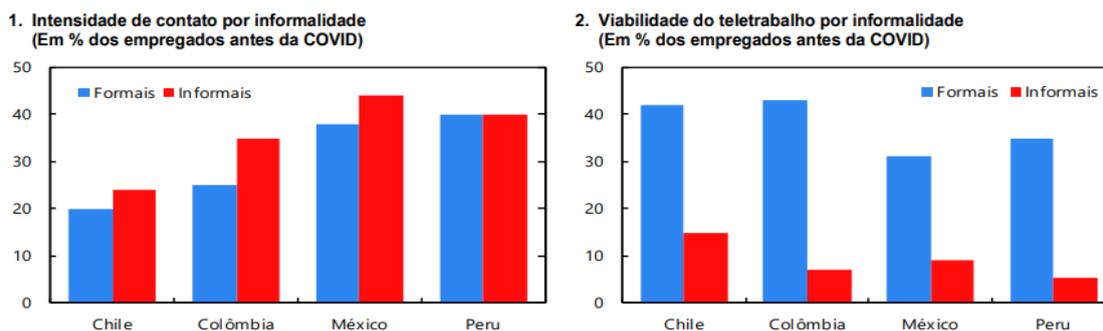
casos novos diários varia muito entre os países. Com esta discrepância do efeito da pandemia em cada país, é difícil perceber um padrão.

Além disso, o que pode influenciar também nas diferenças de impacto do vírus nos países, são os seus respectivos feriados ou datas festivas. Como cada país tem seus próprios feriados, a curva de casos de um país pode se comportar de maneira diferente da curva de casos de outro país em determinada data, uma vez que as curvas podem sofrer influências desses eventos de feriado, onde por exemplo, pode haver menos pessoas nos locais de trabalho e portanto menos risco de transmissão, ou pode ter mais pessoas na rua e portanto maior risco de transmissão.

Um exemplo disso é o dia 07 de setembro de 2020, feriado nacional no Brasil. Nessa data, segundo o report do Google, o Brasil tem uma variação percentual na quantidade de pessoas nos locais de trabalho de -61% em relação a um valor base, ou seja, queda do número de pessoas em locais de trabalho, e uma variação percentual de 3% em parques e praias, ou seja, aumento do número de pessoas em parques e praias. Enquanto que a Argentina, que não tem esse feriado, tem uma variação percentual de -27% na quantidade de pessoas nos locais de trabalho em relação a um valor base, ou seja, queda menor quando comparado ao Brasil, e uma variação percentual de -72% em parques e praias, ou seja, diferente do Brasil o número de pessoas em parques e praias diminuiu.

No quesito laboral, as diferenças ocorrem em diversos aspectos. A figura 8 apresenta as diferenças entre trabalhadores formais e informais em relação à intensidade de contato e viabilidade do teletrabalho em cinco países da América Latina (ALC): Brasil, Chile, Colômbia, México e Peru. À esquerda, a intensidade de contato se refere à frequência e proximidade do contato físico entre trabalhadores e outras pessoas no ambiente de trabalho. À direita, a viabilidade do teletrabalho se refere à capacidade dos trabalhadores de realizar suas atividades remotamente, por meio de tecnologias digitais.

Figura 8 – Teletrabalho e informalidade. Em 1, a intensidade de contato por informalidade e em 2 viabilidade do teletrabalho por informalidade.



Fonte: Fundo Monetário Internacional | Outubro 2020

No entanto, as diferenças entre os países são significativas. Por exemplo, no Brasil e no México, a parcela de trabalhadores informais empregados em funções com contato intensivo é muito maior do que a de trabalhadores formais, enquanto no Chile e na Colômbia, as diferenças são menores. Além disso, a parcela de trabalhadores informais com funções altamente adaptáveis ao teletrabalho é muito menor no Brasil e no México do que nos outros países. No entanto, as diferenças entre os países são significativas. Por exemplo, no Brasil e no México, a parcela de trabalhadores informais empregados em funções com contato intensivo é muito maior do que a de trabalhadores formais, enquanto no Chile e na Colômbia, as diferenças são menores. Além disso, a parcela de trabalhadores informais com funções altamente adaptáveis ao teletrabalho é muito menor no Brasil e no México do que nos outros países. Essas diferenças refletem as características específicas de cada país em termos de estrutura econômica, mercado de trabalho e políticas públicas.

A mineração de dados, aliada à mineração de processos, visa estabelecer visões mais claras acerca de processos complexos tal como a pandemia foi.

3.2. Coleta de dados

Este trabalho é um estudo de caso que identifica as regras implícitas que orientaram as ações governamentais em resposta à pandemia de COVID-19 durante março de 2020 até outubro de 2022, em países da América Latina das sub-regiões América Central e Caribe (Costa Rica, Haiti, Guatemala, República Dominicana, El

Salvador, Nicarágua, Honduras, Panamá), Região Andina (Bolívia, Brasil, Colômbia, Equador, Peru), Cone Sul (Argentina, Chile, Uruguai, Paraguai) e México. A figura 9 destaca quais foram os países analisados:

Figura 9 - Mapa com os países analisados.



Fonte: Elaboração própria

Cuba foi removida da análise, uma vez que não possui dados referentes à mobilidade no país durante a pandemia, como se observou no dataset original.

A Venezuela também foi retirada da análise, dado que o governo venezuelano não publica dados epidemiológicos há vários anos devido ao colapso quase total do sistema de saúde do país. O sistema de vigilância do país não funciona corretamente e a capacidade de diagnóstico é limitada, como explica Cunha *et al.* (2020).

Então, ao todo são 18 países, todos eles pertencentes a América Latina.

A base de dados usada neste trabalho foi gerada a partir da compilação das informações de diversas fontes. Estas informações dizem respeito a:

- i) Dados de mobilidade: são dados de mobilidade obtidos através de um relatório do Google Mobility Data;
- ii) Indicadores socioeconômicos: dados obtidos nos sites do Banco Mundial, da Organização internacional do Trabalho, e Nações Unidas;

- iii) Dados relacionados com a resposta nacional à pandemia. Esses dados foram obtidos no site Oxford Covid Policy Tracker; dados relacionados à curva epidêmica e vacinação, disponíveis na base de dados Our World in Data.

3.3 Pré-processamento

A compilação dessas fontes de dados foi realizada no programa python, assim como o cálculo da média móvel de sete dias dos novos casos diários referente a cada país.

Com a média móvel de sete dias dos novos casos diários, demarcou-se de forma sequencial, utilizando um script do programa R de Jaszbiak (2022), os pontos de inflexões dessas curvas, ou seja, o momento em que a curva de média móvel de 7 dias de casos novos diários muda de inflexão. Isso foi calculado através da segunda derivada da curva de cada país, de maneira que o ponto de inflexão seria determinado no instante em que a segunda derivada muda de sinal. O objetivo foi capturar os momentos nos quais a curva de cada país começa a acelerar ou desacelerar.

Dessa maneira, pôde-se delimitar quais medidas de contenção ou vacinação foram tomadas em período de aceleração ou desaceleração e, assim, caracterizar o possível efeito das medidas governamentais na mudança do comportamento da curva de casos de COVID-19.

Além disso, criaram-se outras variáveis auxiliares associadas a cada país, representando datas de acontecimentos considerados relevantes para a análise. As variáveis e a descrição delas são:

- Marco Vacinação Por Pelo Menos Uma Dose: variável que demarca as datas onde percentual de vacinados por pelo menos uma dose atinge um múltiplo de 10;
- Primeiro Registro de Caso: variável que indica a data do primeiro registro de caso de COVID-19 para cada país;
- Início da aceleração ou desaceleração de casos: variável que indica a data

de início da desaceleração ou aceleração da curva de casos de cada país;

- Data da Primeira Vacina Aplicada: variável que indica a data da primeira vacina aplicada em cada país.

Vale ressaltar que ainda há as variáveis que representam o nível de abrangência (regional ou nacional) de uma dada medida. São elas: C1_Flag, C2_Flag, C3_Flag, C4_Flag, C5_Flag, C6_Flag, C7_Flag, E1_Flag, H1_Flag, H6_Flag, H7_Flag, H8_Flag, no qual as duas primeiras letras representam qual medida aquele flag corresponde. Então, por exemplo, o C4_Flag é o flag de C4_Restrictions on gathering. Os outros flags seguem a mesma lógica. Os dados das medidas de contenção, vacinação, e mobilidade ainda foram ajustados de modo que as informações de cada medida fossem dispostas em linhas e não em colunas, como foram originalmente obtidas.

O dataset antes da formatação pode ser exemplificado como mostra a tabela 1:

Tabela 1 - Dataset antes da formatação

Date	CountryName	...	C1_School closing	...	H8_Protection of elderly people	workplaces_percent_change_from_baseline	...
01/01/2020	Brazil	...	0	...	0	0%	...
...

Fonte: elaboração própria

Depois de formatado, o dataset pode ser exemplificado como mostra a tabela 2. As medidas dispostas anteriormente em colunas, passam a ser dispostas em linha como é possível perceber na coluna “Eventos”:

Tabela 2 - Dataset depois da formatação

Date	CountryName	...	Eventos	Valor	...
01/01/2020	Brazil	...	C1_School closing	0	...
01/01/2020	Brazil	...	H8_Protection of elderly people	0	...
01/01/2020	Brazil	...	workplaces_percent_change_from_baseline	0%	...
...

Fonte: elaboração própria

Essa formatação habilita a leitura e interpretação do dataset nas ferramentas usadas neste trabalho.

Os registros de eventos anteriores ao primeiro caso registrado no país foram removidos. Houve a remoção também dos registros de eventos posteriores a outubro de 2022, dado que é o limite dos dados de mobilidade do Google. Com isso, o período de análise limitou-se de 01/01/2020 a 15/10/2022.

Ainda como pré-processamento, para as variáveis de medidas de mobilidade, as variações tanto positivas quanto negativas foram reclassificadas com objetivo de diminuir a variabilidade dos valores e, com isso, diminuir a complexidade das análises. As variáveis de medidas de mobilidade são:

- retail_and_recreation_percent_change_from_baseline,
- grocery_and_pharmacy_percent_change_from_baseline,
- parks_percent_change_from_baseline,
- transit_stations_percent_change_from_baseline,
- workplaces_percent_change_from_baseline,
- residential_percent_change_from_baseline

A reclassificação foi feita da seguinte maneira:

- I. Quando o módulo do percentual fosse abaixo de 33%: seria classificado como “variação baixa”
- II. Quando o módulo do percentual fosse igual ou maior que 33% e menor que 66% : variação média
- III. Quando o módulo do percentual fosse maior que 66% : variação alta

Portanto, por exemplo, caso o percentual da variável `workplaces_percent_change_from_baseline` fosse de -23% em um dado registro, essa variação seria classificada como uma variação negativa baixa. É uma variação negativa, pois o sinal do registro é negativo. E é uma variação baixa, pois 23% trata-se de um percentual abaixo de 33%. Logo, esse registro seria classificado como: variação negativa baixa de `workplaces_percent_change_from_baseline`

Para as variáveis de medidas farmacêuticas e não farmacêuticas, também foram feitas reclassificações pelo mesmo motivo mencionado para as variáveis de mobilidade: as variáveis apresentavam muitos valores, o que complicaria mais a mineração, então foi feita a redução do número de distinções que as variáveis podem ter, no que se refere aos seus valores. As reclassificações foram baseadas no nível de rigidez da medida, de modo que a medida apenas tivesse, na maioria dos casos, duas classificações: uma que indicava uma rigidez fraca da medida e outra que indicasse uma rigidez forte da medida. Então, por exemplo, a medida de contenção ‘C1_School closing’ tinha 3 valores (1,2,3) que variam conforme grau de rigidez da medida no momento do registro, mas após a reclassificação a medida ficou com apenas 2 valores, um representando uma rigidez mais forte da medida e a outra representando uma rigidez mais leve, sendo o valor que apresentasse a rigidez máxima fosse isolado, de modo a representar grau de rigidez forte e os valores restantes (1,2) foram reclassificados para um valor único que representasse uma rigidez fraca. Além da variável ‘C1_School closing’, outras variáveis foram tratadas dessa maneira. São elas : 'C2_Workplace closing', 'C4_Restrictions on gatherings', 'C6_Stay at home requirements', 'C8_International travel controls', 'H6_Facial Coverings', 'H8_Protection of elderly people’.

Outra tratativa foi a remoção das variáveis onde mais de 20% dos seus registros eram valores nulos. A definição do número percentual de 20% foi escolhido de maneira arbitrária. As variáveis eliminadas foram:

- 'E3_Fiscal measures',
- 'E4_International support',
- 'H4_Emergency investment in healthcare',
- 'V2B_Vaccine age eligibility/availability age floor (general population summary)',
- 'V2C_Vaccine age eligibility/availability age floor (at risk summary)',
- 'V2D_Medically/ clinically vulnerable (Non-elderly)',
- 'V2E_Education',
- 'V2F_Frontline workers (non healthcare)',
- 'V2G_Frontline workers (healthcare)',
- 'V4_Mandatory Vaccination (summary),
- Marco Vacinação Por Pelo Menos Uma Dose Primeiro Registro de Caso,
- Início da 1° desaceleração ,
- Início da 1° aceleração ,
- Início da 2° desaceleração ,
- Início da 2° aceleração ,
- Início da 3° desaceleração ,
- Início da 3° aceleração ,
- Início da 4° desaceleração ,
- Início da 4° aceleração ,
- Data Primeira Vacina Aplicada

Foram também removidas variáveis altamente correlacionadas, de modo a se fazer uma seleção de variáveis e diminuir a complexidade da análise. Nesse sentido, utilizou-se o fator de inflação da variância (VIF), que mede a multicolinearidade, isto é, mede o quanto a variável é correlacionada com um ou mais variáveis. O cálculo é feito em vários passos. A cada passo, o VIF é calculado, identificando aquela com o maior VIF. Essa variável é então removida do modelo, e os VIFs das variáveis restantes são atualizados. O processo é repetido até que todas as variáveis no modelo tenham VIFs aceitáveis, indicando baixa correlação entre elas. No caso, foram removidas variáveis com $VIF > 10$:

- hdi,
- 'StringencyIndex', 'pct_informal_tot_jobs',
- 'gdp_per_capita',
- 'people_at_least_first_dose_vaccinated_per_population',
- 'gini',
- 'mys',
- 'retail_and_recreation_percent_change_from_baseline',
- 'H1_Public information campaigns',
- 'V1_Vaccine Prioritisation (summary)',
- 'people_at_least_first_dose_vaccinated_per_population',
- 'H7_Vaccination policy',
- 'C8_International travel controls_Agrup',
- 'C2_Workplace closing_Agrup',
- 'transit_stations_percent_change_from_baseline',
- 'H2_Testing policy',
- 'C3_Cancel public events',
- 'H6_Facial Coverings_Agrup',
- 'V2A_Vaccine Availability (summary)',
- 'C1_School closing_Agrup

Restaram, portanto, 13 variáveis de evento. São elas:

- 'C4_Restrictions on gathering',
- 'C5_Close public transport',
- 'C6_Stay at home requirements',
- 'C7_Restrictions on internal movement',
- 'E1_Income support',
- 'E2_Debt/contract relief',
- 'H3_Contact tracing',
- 'H5_Investment in vaccines',
- 'H8_Protection of elderly people',

- 'Vaccine Financial Support (summary)',
- 'parks_percent_change_from_baseline',
- 'workplaces_percent_change_from_baseline',
- 'residential_percent_change_from_baseline'.

Com esses eventos, foi realizada outra tratativa: o nome do evento e o seu respectivo valor foram concatenados, entendendo que cada valor de uma determinada variável representa um evento diferente. Logo, o dataset contendo as colunas “Eventos” e “Valor” passou a ter uma coluna de “Eventos”, cujo qual é a concatenação dos valores da coluna “Eventos” com os valores da coluna “Valor”. A tabela 3 exemplifica como ficou o dataset.

Tabela 3 - Dataset depois da concatenação das colunas “Eventos” e “Valor”.

Date	CountryName	...	Eventos	...
01/01/2020	Brazil	...	C1_School closing_0	...
01/01/2020	Brazil	...	H8_Protection of elderly people_0	...
01/01/2020	Brazil	...	workplaces_percent_change_from_ baseline_0%	...
...

Fonte: elaboração própria

A figura 10 ilustra como as variáveis foram nomeadas e ao lado o que essa variável representa:

Figura 10 - Variáveis de eventos e o significado de cada uma delas.

Fonte	Medida	Significado
OxFord Policy Tracker	Restrictions on gatherings	
	C4_Restrictions on gatherings_Agrup_0	No Restrictions
	C4_Restrictions on gatherings_Agrup_1	Restrição em grandes aglomerações (com mais de 1000 pessoas)
	C4_Restrictions on gatherings_Agrup_4	Restrição em pequenas aglomerações (com menos de 1000 pessoas)
	Closing of public transport	
	C5_Closing of public transport_1	Recomenda-se o fechamento ou redução significativa do transporte público
	C5_Closing of public transport_2	O fechamento do transporte público é obrigatório
	Stay at home requirements	
	C6_Stay at home requirements_Agrup_2	Restrição suaves em não sair de casa
	C6_Stay at home requirements_Agrup_3	Restrição severas em não sair de casa
	restrictions on Internal movement	
	C7_restrictions on Internal movement_1	Recomenda-se evitar viagens entre regiões/cidades
	C7_restrictions on Internal movement_2	Há restrições de movimentação interna em vigor
	Income support	
	E1_Income support_1	O governo está substituindo menos de 50% do salário perdido
	E1_Income support_2	O governo está substituindo 50% ou mais do salário perdido.
	debt/contract relief for households	
	E2_debt/contract relief for households_1	Alívio restrito (específico para um tipo de contrato)
	E2_debt/contract relief for households_2	Ampla alívio de dívidas/contratos
	Contact tracing	
	H3_Contact tracing_0	Sem rastreamento de contatos
	H3_Contact tracing_1	Rastreamento de contatos limitado; não realizado para todos os casos
	H3_Contact tracing_2	Rastreamento abrangente de contatos; realizado para todos os casos identificados
	Investment in vaccines	
	H5_Investment in vaccines	Registro de valor monetário em USD
Protection of elderly people		
H8_Protection of elderly people_2	Restrições parcas para isolamento e higiene em Lares de Idosos, algumas limitações para visitantes externos e/ou restrições para proteção de idosos em casa	
H8_Protection of elderly people_3	Restrições severas para isolamento e higiene em Lares de Idosos, proibição de todos os visitantes externos não essenciais, e/ou obrigatoriedade para todos os idosos ficarem em casa, não saindo com exceções mínimas, e sem receber visitantes externos	
Vaccine Financial Support (summary)		
V3_Vaccine Financial Support (summary)_1	custo total para o indivíduo para todas as categorias identificadas em V2	
V3_Vaccine Financial Support (summary)_2	custo total para o indivíduo para algumas categorias identificadas em V2, algum subsídio para outras categorias	
V3_Vaccine Financial Support (summary)_3	algum financiamento parcial pelo governo para todas as categorias identificadas em V2	
V3_Vaccine Financial Support (summary)_4	financiamento parcial pelo governo para algumas categorias identificadas em V2, financiamento total para outras categorias	
V3_Vaccine Financial Support (summary)_5	todas as categorias totalmente financiadas pelo governo	
Google Mobility Data	Mobilidade	
	retail_and_recreation_percent_change_from_baseline	Mudança percentual nas visitas a restaurantes, cafés, centros comerciais, museus, bibliotecas e cinemas em comparação com a linha de base
	grocery_and_pharmacy_percent_change_from_baseline	Mudança percentual nas visitas a lugares como supermercados, armazéns de alimentos, farmácias e drogarias em comparação com a linha de base
	parks_percent_change_from_baseline	Mudança percentual nas visitas a locais como parques e praias em comparação com a linha de base
	transit_stations_percent_change_from_baseline	Mudança percentual nas visitas a paradas de transporte público em comparação com a linha de base
	workplaces_percent_change_from_baseline	Mudança percentual nas visitas aos locais de trabalho em comparação com a linha de base
	residential_percent_change_from_baseline	Mudança percentual nas visitas a áreas residenciais em comparação com a linha de base

Fonte: Elaboração própria

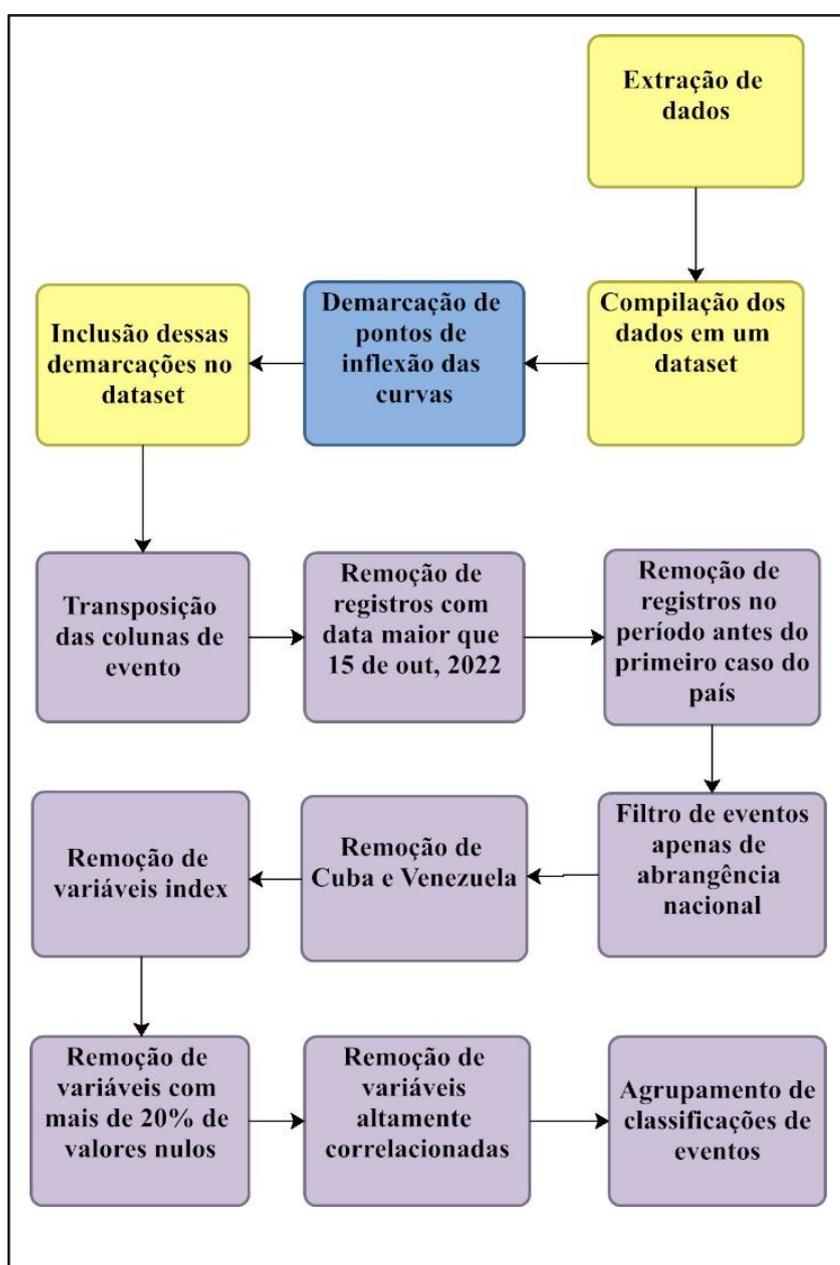
Vale ressaltar que só foram pegos os registros das informações de abrangência nacional, ou seja, intervenções/medidas válidas para o país inteiro. Os campos que representam os níveis de abrangência das medidas são: C4_Flag, C5_Flag, C6_Flag, C7_Flag, E1_Flag, H8_Flag. Os registros nos quais o valor do flag é 0, foram removidos, dado que representam registros das intervenções/medidas regionais, ou seja, medidas dirigidas a apenas algumas regiões de um determinado país. Isso porque esta

é uma análise referente aos países como um todo. Então, foram considerados apenas os registros das intervenções/medidas nacionais(valor do flag é 1) , ou seja, medidas dirigidas a todo país.

Além disso, foram removidas as variáveis que representavam índices das variáveis das medidas de contenção. Essas variáveis foram removidas, pois eram redundantes.

A figura 11 resume as etapas realizadas no pré-processamento do dataset:

Figura 11 - Etapas de extração e pré-processamento.



Fonte: Elaboração própria

Em amarelo, as etapas feitas no programa python, em azul as etapas feitas no programa R e em roxo as etapas feitas no SQL Server.

As variáveis do dataset resultante, juntamente com as suas descrições estão disponíveis no link: <https://github.com/natalia-sendra/Covid-Mining-Process-Latin-America/blob/main/DatasetResultanteDescricao.xlsx>)

3.4. Formatação do log de eventos

Para a mineração de processos, faz-se necessário a criação do log de eventos. Para tanto, foi estabelecido que a instância do processo é composta pelo país juntamente com número que obedece a ordem pela qual determinada inflexão da curva de casos acontece. Então, por exemplo, para a segunda inflexão da curva de casos do Chile, a instância é ChileT2. O sufixo T2 é uma janela temporal de análise definida a partir da identificação de pontos de mudança de inflexão na série temporal da variável de “média móvel de 7 dias de novos casos diários”. A data do início da janela, portanto, é a data do ponto de inflexão e a data final é a data do ponto de inflexão que vem logo em seguida. Então, para um dado país, cada janela de análise contém os eventos que ocorreram entre dois pontos de mudança consecutivos na série temporal da variável de classe. O objetivo foi capturar os momentos nos quais a curva de cada país começa a acelerar ou desacelerar. Dessa maneira, de acordo com Jazbik (2022), ao dividir o processo em janelas de tempo, é possível identificar mudanças significativas em diferentes etapas do processo e entender como essas mudanças afetam o resultado final. Além disso, a análise de janelas de tempo pode ajudar a identificar padrões e tendências no processo, bem como detectar desvios e anomalias que possam indicar problemas ou oportunidades de melhoria, como discorre Jazbik (2022).

Os eventos do processo são as medidas de contenção e eventos de mobilidade. O `start_timestamp` é a data de quando o evento começa e o `end_timestamp` é a data de quando esse mesmo evento termina.

Abaixo uma amostra do log de eventos:

Figura 12 - Log de eventos referente a eventos e medidas de contenção da COVID-19.

Instancia	NomeEvento	start_timestamp	end_timestamp
ArgentinaT1	C6_Stay at home requirements_Agrup_2	2020-04-27 00:00:00.000	2020-08-24 00:00:00.000
ArgentinaT1	C6_Stay at home requirements_Agrup_3	2020-03-19 00:00:00.000	2020-04-26 00:00:00.000
ArgentinaT1	C7_Restrictions on internal movement_2	2020-05-11 00:00:00.000	2020-08-24 00:00:00.000

Fonte: Elaboração própria

Cada linha refere-se a um registro de evento de uma dada instância. O log de eventos conta com: um identificador de ocorrência definido pela coluna “Instancia”, nome do evento definido pela coluna “Nomeevento” e data do evento demarcado por “start_timestamp” e “end_timestamp”.

3.5. Descoberta do processo de evolução da pandemia de COVID-19 utilizando a abordagem declarativa

Uma das abordagens da mineração de processos é a descoberta, utilizada com intuito de se encontrar a modelagem que melhor descreve o processo.

3.5.1 Modelagem declarativa

Como trata-se de um processo muito complexo e diverso, foram analisadas as regras com support e confidence mais altos, apresentadas na Tabela 4.

Percebe-se que a primeira regra da Tabela 4 indica que, quando há uma redução nas atividades nos locais de trabalho, é altamente provável (Suporte de 95,79% e confiança de 89,05%) que haja um aumento nas atividades residenciais durante os períodos de aceleração da curva de casos, o que pode indicar uma mudança comportamental significativa da população. A terceira regra complementa essa ideia.

A segunda regra mostra que a redução inicial nas atividades nos locais de trabalho (baixa variação negativa) é seguida por uma diminuição mais significativa (média variação negativa), o que pode significar que a redução de atividades nos locais de trabalho ocorreu de maneira gradual.

A terceira regra está relacionada com a situação das regras anteriores. Quando há uma média variação negativa nos locais de trabalho (workplaces_percent_change_from_baseline), então precede-se deste evento a medida de contenção que estabelece que a população não deve sair de casa salvo mínimas exceções.

A quarta regra está relacionada com a situação das regras anteriores. Quando há uma média variação negativa nos locais de trabalho (workplaces_percent_change_from_baseline), então estarão em vigor restrições internas de movimento entre cidade e regiões (C7_Restrictions on internal movement_2).

Tabela 4 - Principais regras do momento de aceleração

Id	Template	TaskA	TaskB	Support	Confidence
1	Response	workplaces_percent_change_from_baseline_Média variação negativa	residential_percent_change_from_baseline_Baixa variação positiva	95,79%	89,05%
2	Precedence	workplaces_percent_change_from_baseline_Baixa variação negativa	workplaces_percent_change_from_baseline_Média variação negativa	88,32%	82,10%
3	Precedence	C6_Stay at home requirements_Agrup_2	workplaces_percent_change_from_baseline_Média variação negativa	71,96%	66,89%

4	Precedence	C7_Restrictions on internal movement_2	workplaces_percent_change_from_baseline_Média variação negativa	71,50%	66,46%
---	------------	--	--	--------	--------

Fonte: Elaboração própria

Enquanto que para os momentos de desaceleração, como pode ser visto na tabela 5 abaixo, os regras de id 1, 3, e 4 indicam uma associação positiva entre o suporte financeiro e a redução nas atividades nos locais de trabalho durante momentos de desaceleração. Então, o suporte financeiro está consistentemente associado a uma diminuição nas atividades dos locais de trabalho. Como esta associação entre o suporte financeiro e a diminuição das atividades em determinados locais ocorre de uma maneira tão consistente no momento da desaceleração, pode ser que o suporte financeiro seja um fator relevante para a diminuição das atividades em determinados locais.

A regra de id 2 sugere que uma diminuição baixa das atividades em parques durante períodos de desaceleração é precedida pela medida de contenção que estabelece que a população não deve sair de casa salvo mínimas exceções.

A regra de id 5 sugere que a proteção de idosos está associada a uma diminuição das atividades em parques durante períodos de desaceleração.

Tabela 5 - Principais regras do momento de desaceleração

Id	Template	TaskA	TaskB	Support	Confidence
1	Precedence	E1_Income support_1	workplaces_percent_change_from_baseline_Baixa variação negativa	80,42%	68,05%
2	Precedence	C6_Stay at home requirements_Agrup_2	workplaces_percent_change_from_baseline_Baixa variação negativa	78,90%	66,76%

3	Alternative Response	E1_Income support_1	workplaces_percent_change_from_baseline_Baixa variação negativa	83,33%	56,41%
4	Responded Existence	E1_Income support_1	workplaces_percent_change_from_baseline_Baixa variação negativa	81,71%	55,31%
5	Responded Existence	H8_Protection of elderly people_Agrup_2	parks_percent_change_from_baseline_Média variação negativa	67,44%	49,80%

Fonte: Elaboração própria

As regras apresentadas foram descobertas usando todo o dataset, com suporte e confiança altos. Por isso, tais regras refletem o que aconteceu no geral entre todos os países, ou seja, as similaridades de comportamento entre eles, confirmando conhecimento que já era conhecido ou esperado com relação às medidas de contenção. Adicionalmente, fez-se necessário descobrir as diferenças de regras entre países, que expliquem as diferenças da evolução da pandemia entre eles.

Nesse sentido, uma análise mais individual dos países revela um impacto expressivo da pandemia nos indicadores socioeconômicos de vários países latino-americanos. Países como El Salvador, México, Uruguai e Chile experimentaram longos períodos (278, 267, 267 e 253 dias, respectivamente) com uma variação negativa significativa no número de pessoas em locais de trabalho.

Destaca-se também que, em 2020, Chile, México e Peru testemunharam os maiores decréscimos no emprego informal em áreas urbanas, conforme explica Acevedo (2021). Esse fenômeno pode ser interpretado como uma resposta adaptativa à crise, indicando mudanças substanciais nas dinâmicas do mercado de trabalho informal.

Além disso, há uma alta variação negativa em parques, isto é, uma queda alta na quantidade de pessoas comparado a um valor base, observada no Panamá, Uruguai e Argentina ao longo de 187, 147 e 126 dias, respectivamente. Esses países são os que tiveram maior tempo dessa variável, o que reflete o impacto direto das medidas de contenção da pandemia no setor de lazer e recreação. A sincronia entre a aceleração desses declínios e a propagação do vírus sugere uma conexão intrínseca entre as restrições implementadas e as mudanças nos padrões de comportamento social. Em contrapartida, países como Argentina, Bolívia, Peru, Equador, Colômbia e Paraguai, que adotaram medidas rigorosas em resposta à pandemia, destacam-se por sua abordagem proativa na gestão da crise, apesar dos desafios econômicos associados (ACEVEDO *et al.*, 2021).

Ainda na análise a nível de país, utilizou-se também as regras do Declare Process Mining listadas no trabalho para entender como cada país se comportou. Para tanto, utilizou-se a variável ‘workplaces_percent_change_from_baseline_Alta variação negativa’ para entender as relações dessa variável em diferentes países. A tabela 6 mostra o resultado relacionados a essa variável com 5 maiores confiança e support.

Tabela 6 - Regras alta confiança por país

Id	Template	TaskA	TaskB	Support	Confidence	Country
1	response	C4_Restric tions on gatherings_ Agrup_4	workplaces_perce nt_change_from_ baseline_Alta variação negativa	100,00%	75,00%	Paraguay
2	Alternative Response	C4_Restric tions on gatherings_ Agrup_4	workplaces_perce nt_change_from_ baseline_Alta variação negativa	100,00%	75,00%	Paraguay
3	alternate Precedence	C4_Restric tions on	workplaces_perce nt_change_from_	100,00%	75,00%	Paraguay

		gatherings_	baseline_Alta			
		Agrup_4	variação negativa			
4	alternate	C6_Stay at	workplaces_perce	100,00%	75,00%	Paraguay
	Precedence	home	nt_change_from_			
		requiremen	baseline_Alta			
		ts_Agrup_2	variação negativa			
5	Precedence	C6_Stay at	workplaces_perce	100,00%	75,00%	Paraguay
		home	nt_change_from_			
		requiremen	baseline_Alta			
		ts_Agrup_2	variação negativa			

Fonte: Elaboração própria

A alta queda no número de pessoas nos locais de trabalho parece estar mais relacionada a restrições a aglomerações, requisitos de permanência em casa, restrições ao movimento entre as cidades/regiões. O fato interessante reside do Paraguai aparecer repetidamente por obter as maiores confianças e suportes no que se refere às regras relacionadas a variável ‘workplaces_percent_change_from_baseline_Alta variação negativa’. Segundo dados do fundo monetário internacional¹¹, o Paraguai teve sucesso em conter a pandemia logo no seu início com poucos casos e vítimas.

Com objetivo de se comparar o comportamento entre os países e analisar a discrepância entre os países, utilizou-se novamente a variável ‘workplaces_percent_change_from_baseline_Alta variação negativa’ e com as mesmas regras, no entanto, dessa vez analisando os resultados relacionados a essa variável com as menores confianças e suportes.

Tabela 7 - Regras baixa confiança por país

Id	Template	TaskA	TaskB	Support	Confidence	Country
1	Alternative Response	C4_Restrictions on gatherings_Agrup_4	workplaces_perce nt_change_from_ baseline_Alta variação negativa	20,00%	20,00%	Haiti
2	response	C4_Restrictions on gatherings_Agrup_4	workplaces_perce nt_change_from_ baseline_Alta variação negativa	20,00%	20,00%	Haiti
3	alternate Precedence	C4_Restrictions on gatherings_Agrup_4	workplaces_perce nt_change_from_ baseline_Alta variação negativa	14,29%	7,14%	Bolívia
4	alternate Precedence	C6_Stay at home requirements_Agrup_2	workplaces_perce nt_change_from_ baseline_Alta variação negativa	14,29%	7,14%	Peru
5	Precedence	C6_Stay at home requirements_Agrup_2	workplaces_perce nt_change_from_ baseline_Alta variação negativa	100,00%	75,00%	Haiti

Fonte: Elaboração própria

No Haiti, a relação entre restrições a aglomerações com uma queda significativa na quantidade de pessoas nos locais de trabalho, ocorre com um suporte de 20% e uma confiança de 20%. Isso tanto para as regras de Response quanto para Alternative Response. Na Bolívia, essa relação ocorre com um suporte de 14,29% e uma confiança

de 7,14% . A hipótese levantada por esta regra é que a medida de contenção “restrições a aglomerações” nesses países não foi obedecida suficientemente, dado que não houve uma confiança maior na “diminuição da quantidade de pessoas nos locais de trabalho” em resposta a essa medida.

Peru tem uma confiança pequena na relação entre ‘requisitos de permanência em casa’ com a queda significativa na quantidade de pessoas nos locais de trabalho. Trata-se de um país com altos índices de empregos formais e informais nos quais há uma alta intensidade de contato, conforme visto no capítulo anterior. Além disso, o emprego informal possui uma representatividade maior em relação ao emprego formal.

Comparando com o Paraguai, esses países passaram mais tempo com uma baixa significativa de pessoas nos locais de trabalho. No entanto, segundo o Índice Global Health Security Index (GHS) de 2021, que avalia as capacidades de 195 países para se prepararem para epidemias e pandemias, o Paraguai é o segundo mais bem ranqueado nesse índice entre esses países, só atrás do Peru.

Tabela 8 - Duração média de dias da ‘variável workplaces_percent_change_from_baseline_Alta variação negativa’

País	Duração média de dias
Paraguay	10
Peru	18
Bolívia	20
Haiti	22

Fonte: Elaboração própria

3.5.2 Modelagem imperativa

Além do estudo de caso principal com a análise aprofundada com a abordagem declarativa, também foi utilizada uma ferramenta que adota o paradigma imperativo, Celonis, para ilustrar como a visualização dessa abordagem é mais complexa para o estudo.

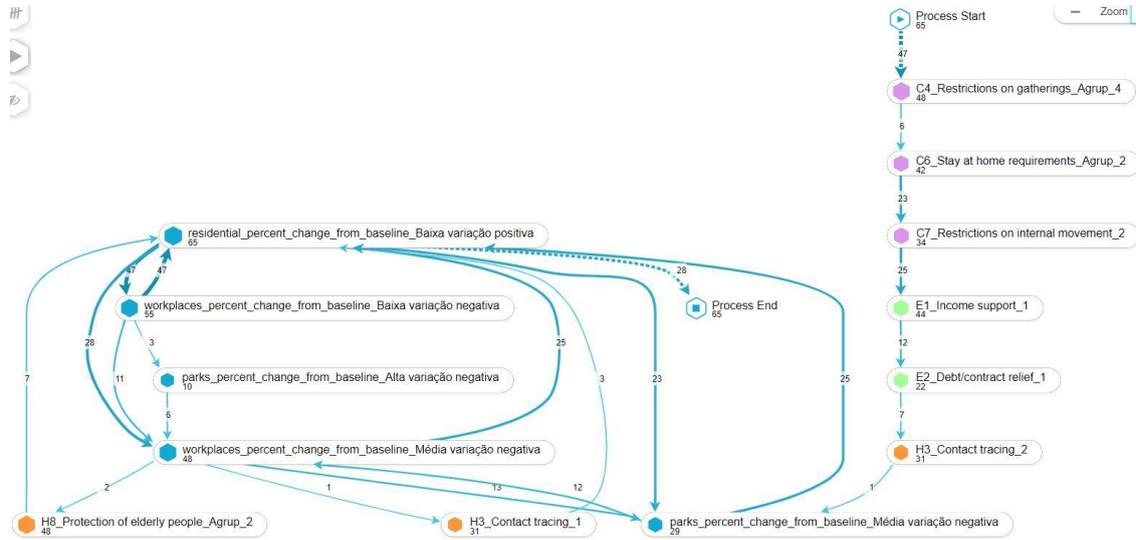
As etapas para utilizar a ferramenta Celonis foram baixar o arquivo Excel com a base de dados do estudo de caso, determinar qual o tipo de dado de cada coluna, como string(texto), integer(número inteiro) etc, selecionar as colunas correspondentes a cada parte do processo, como qual é a coluna da instância, da data, do nome do evento, o momento, e após o grafo do processo estar pronto ajustar o quanto seria observado.

As colunas a serem selecionadas são case ids (instância), activities (nomeevento), timestamp (start_date ou end_date) e uma opcional de sorting information (momento). Utilizou-se start_date na coluna de data, filtrada por aceleração e posteriormente desaceleração. A ferramenta oferece espaço para anotações de insights (percepções) e como já comentado, ajustes em relação à quantidade de atividades e conexões entre elas. O nível de complexidade da análise pode ser muito grande e esses ajustes podem tornar mais visível a parte do processo que mais interessar.

Para o entendimento do processo como um todo, se tiver muitas atividades e interconexões, como é o caso, ele vira um diagrama espaguete. É viável identificar os eventos de início e término, uma vez que se manifestam por meio de linhas distintas. Contudo, no decorrer do processo, depara-se com uma considerável sobreposição de linhas, o que impõe restrições substanciais à análise.

A figura 13 representa o processo analisado, com a coluna data de início filtrada por momentos de desaceleração, com 67,3% das atividades (ou seja, das apenas 67,3% foram selecionadas) e as conexões ajustadas para 26,8% (ou seja, das conexões entre todas as atividades apenas 26,8% foram selecionadas) para auxiliar na visualização. Esse ajuste específico de atividades e conexões foi pelo Happy Path (caminho feliz) que é a funcionalidade do Celonis que mostra a atividade inicial mais frequente, a atividade final mais frequente e a variante de processo mais frequente que conecta essas duas atividades.

Figura 13 - Diagrama Processo Celonis com data de início e ajustes



Fonte: Elaboração própria

O processo foi observado também filtrando momentos de aceleração e desaceleração para tentar analisar como as atividades impactam a evolução da COVID-19. As linhas representam as conexões entre as atividades e podem estar simbolizando 5 diferentes conexões dependendo da escolha, pode ser frequência da atividade, do caso, e 3 tipos de tempos de processamento. E, quanto mais grossa a linha maior essa conexão. No caso utilizamos a frequência dos casos, pois auxilia na observação dos momentos de aceleração e desaceleração já que os casos estão subdivididos por seus países e momentos da pandemia. Algumas relações observadas são consideradas como conexões "evidentes". Por exemplo, durante períodos de aceleração da pandemia, foi observada uma baixa variação positiva nos percentuais residenciais em todos os casos. No entanto, é interessante observar que o mesmo ocorreu nas desacelerações.

As principais conexões observadas na aceleração estão representadas na tabela 9.

Tabela 9 - Conexões do momento de aceleração, onde o processo segue do Evento 1 para o Evento 2

Evento 1	Evento 2	Percentual de casos
H8_Protection of elderly people_Agrup_2	V3_Vaccine Financial Support(summary)_5	35%
E1_Income support-_0	E2_Debt/contract relief_0	48%
H3_Contact tracing_0	H8_Protection of elderly people_Agrup_2	11%
H3_Contact tracing_1	H8_Protection of elderly people_Agrup_2	20 %
H3_Contact tracing_2	H8_Protection of elderly people_Agrup_2	21%

Fonte: elaboração própria

Logo, se observa que todos os rastreamentos de contatos precedem em uma quantidade razoável de casos a proteção dos idosos. O agrupamento 2 de proteção dos idosos tem 52% dos casos com algum rastreamento de contatos como precedente. Além disso, se observa em 35% dos casos que a proteção dos idosos com restrição severa vem antes do financiamento de suporte vacinal 5. Por fim, em 48% dos casos nenhum tipo de suporte financeiro acontece antes de não ter nenhum alívio de dívida/contrato.

Já as principais conexões observadas da desaceleração estão representadas na tabela 10:

Tabela 10 - Conexões do momento de desaceleração, onde o processo segue do Evento 1 para o Evento 2

Evento 1	Evento 2	Percentual de casos
H3_Contact tracing_0	H8_Protection of elderly people_Agrup_2	6%
H3_Contact tracing_1	H8_Protection of elderly people_Agrup_2	28%
C7_Restrictions on internal movement_2	E1_Income support_1	38%

Fonte: elaboração própria

Desse modo, assim como na aceleração, se observa em 41% dos casos que a proteção dos idosos com restrição severa vem antes do financiamento de suporte vacinal 5. E, a restrição severa de proteção dos idosos tem 54% dos casos com algum rastreamento de contatos como precedente. No entanto, algo diferente observado foram restrições de movimento interno precedendo o suporte financeiro de menos de 50% do salário.

O que se observa é um aumento de proteção aos idosos na desaceleração, o que impactou positivamente, porém não foi uma mudança tão significativa de suporte. Por fim, foram poucos os insights possíveis através dessa modelagem, pois a falta do paralelismo das datas devido a ferramenta e o tipo de abordagem que tenta restringir um cenário complexo em um único fluxo de processo não levou a uma boa visualização dos dados e a análise exclusiva de relações de precedência que limitam as possíveis conclusões do processo.

4 Resultados

O Celonis mostrou-se uma ferramenta de fácil inserção de dados e geração de análises de processos. No entanto, sua limitação crítica para este caso específico foi a quantidade restrita de colunas com informações disponíveis para seleção. Isso resultou na capacidade de escolher apenas uma coluna para a data das atividades, o que restringiu a análise do processo real. Várias atividades ocorreram ao longo de dias, e não em um único dia.

Portanto, em vez de ter uma análise que incluía datas de início e término, permitindo a observação da ocorrência simultânea das atividades, foram realizadas duas análises separadas, uma apenas com os momentos de aceleração e outra com os de desaceleração. O resultado são dois gráficos do processo que falham em representar as verdadeiras características do processo, incluindo a simultaneidade real das atividades, e não fornecem a duração exata de cada atividade.

Além disso, existe uma limitação no tamanho do log de eventos na versão acadêmica da ferramenta, que foi a usada neste estudo. O log de eventos não pode ultrapassar 100 mil linhas. Então, esta versão da ferramenta do Celonis tem uma limitação de configuração de parâmetros e tamanho do log.

Já a linguagem SQL, utilizada para a obtenção das restrições da modelagem declarativa, permite manipular um grande conjunto de dados, uma vez que é uma linguagem otimizada para operar em grandes conjuntos de dados, o que é crucial ao lidar com logs extensos de eventos, como por exemplo um log de eventos de um período de tempo maior, ou com mais atividades, ou logs de eventos relacionados à evolução da pandemia de todos os países do mundo, não só da América Latina. A eficiência do SQL permite análises robustas e abrangentes. As consultas com a linguagem foram relativamente rápidas, não demorando mais que 1 minuto para a obtenção dos resultados.

Outra vantagem da abordagem declarativa quando implementada no SQL, é a facilidade de manutenção e reprodução do estudo. A clareza na formulação das regras e consultas torna mais simples para outros pesquisadores entenderem, replicarem e estenderem a análise. O Anexo 1 contém os comandos em SQL para cada template de

regra declarativa. Porém, é necessário um conhecimento, ainda que básico, da linguagem SQL para que se consiga atingir os objetivos de pesquisa.

Tabela 11 - Quadro comparativo das ferramentas

Característica	Modelagem Declarativa pelo SQL	Modelagem Imperativa pelo Celonis
Tamanho da base de dados	Capaz de lidar com uma grande capacidade de dados	limite de 100 mil linhas (versão acadêmica)
Adaptabilidade	Alta adaptabilidade a diferentes contextos, com diferentes condições e associações	Pode ser adaptada, mas pode exigir alterações extensivas. Restrita a mostrar sequência de eventos
Curva de Aprendizado	Baixa, caso o usuário esteja familiarizado com SQL	Alta, documentação extensa.
Reusabilidade de Código	Boa reusabilidade, fácil adaptação a diferentes conjuntos de dados, mais fácil de manter	Reusabilidade possível, mas pode ser menos direta
Complexidade de Implementação	Mais fácil de definir as parametrizações	Limitação na coluna de data das atividades (não representa atividades não atômicas)

Fonte: Elaboração própria

Essa tabela resume as principais características consideradas ao comparar as abordagens de modelagem declarativa do SQL e imperativa do Celonis em process mining. No contexto do presente estudo, no qual o log de eventos é complexo e o

comportamento das atividades é altamente variável, a abordagem declarativa destaca-se pela sua facilidade e adequabilidade. Sua expressividade e capacidade de adaptação intuitiva a diferentes contextos a tornam eficaz na análise de logs extensos e intrincados, proporcionando uma abordagem mais acessível e flexível em cenários de grande complexidade e variação nos dados.

Os resultados obtidos por meio da abordagem declarativa, baseada na mineração de processos declarativos, revelaram associações entre atividades e a evolução dos casos da COVID-19. No entanto, não foi possível observar regras que explicassem como diferentes ações tomadas pelos países resultaram em impactos distintos.

No entanto, foram extraídas algumas regras para todos os países: As regras identificadas (ID 1, 3 e 4) na desaceleração evidenciam uma associação consistente entre suporte financeiro e a redução nas atividades nos locais de trabalho, o que sugere que o suporte financeiro desempenha um papel relevante na diminuição das atividades em determinados locais durante momentos críticos.

A regra de id 5 indica que a proteção de idosos está associada a uma diminuição das atividades em parques durante períodos de desaceleração.

Os resultados obtidos alinham-se com os princípios da mineração de processos declarativa, que permite a representação do processo como um conjunto de restrições, o que é essencial para compreender como as atividades individuais influenciam o comportamento global do processo.

Em resumo, a análise declarativa revela padrões significativos nas relações entre atividades, destacando a importância de medidas como suporte financeiro e proteção de idosos em momentos cruciais durante a pandemia, onde há a redução de atividades em locais de trabalho. A compreensão da dinâmica evidenciada por essas regras orienta a implementação eficaz de medidas de contenção durante a pandemia. Observando a redução nas atividades nos locais de trabalho, autoridades podem antecipar aumentos nas atividades residenciais, adaptando estratégias preventivas. Essas regras destacam não apenas restrições nos locais de trabalho, mas também atividades residenciais ao formular políticas de saúde pública em momentos crítico.

5 Conclusão

Este trabalho centrou-se na análise das medidas farmacêuticas e não farmacêuticas nos impactos da pandemia da COVID-19, utilizando técnicas de mineração de processos. Com dados oficiais reportados até 2022, evidenciou-se a necessidade de uma abordagem analítica para compreender os eventos relacionados à evolução da pandemia na América latina.

O objetivo deste estudo foi analisar como as medidas farmacêuticas e não farmacêuticas impactaram a evolução da pandemia da COVID-19. Diante da lacuna de estudos específicos para a América Latina, a escolha da abordagem declarativa de process mining, implementada no SQL, surgiu como uma resposta para explorar padrões e relações, diante do log complexo e altamente variável, revelado nesse cenário.

A coleta de dados envolveu compilação de dados de fontes como Oxford Covid Policy Tracker, Google Mobility data e The World Bank, através da linguagem python. Para efeitos de análises, foram demarcadas as inflexões da curva a média móvel de sete dias dos novos casos diários de cada país, e assim, foram obtidos os momentos de aceleração e desaceleração da curva. Esse processo permitiu caracterizar o impacto das medidas governamentais nos diferentes contextos, tanto em momentos de aceleração como desaceleração da curva de casos. Foi utilizada a mineração de processos para obter os apontamentos do estudo.

A abordagem declarativa, ao utilizar mineração de processos, destacou padrões consistentes, como a relação entre suporte financeiro e a redução nas atividades nos locais de trabalho durante momentos críticos. Além disso, a proteção de idosos mostrou-se associada à redução de atividades em parques durante períodos de desaceleração, o que pode provocar a queda de casos.

Contribuindo para a área de pesquisa, este estudo ressalta a eficácia da abordagem declarativa, especialmente em contextos complexos e variáveis. As análises revelam insights para a tomada de decisões estratégicas e a formulação de políticas de saúde pública em outras pandemias que porventura podem acontecer futuramente.

Como propostas de melhorias, sugere-se explorar a abordagem object-centric process mining (mineração de processos orientada ao objeto) em trabalhos futuros, enriquecendo a análise com uma perspectiva mais orientada ao objeto e uma possível

abordagem híbrida, que como pontuado no texto necessita ainda de mais estudos e desenvolvimento antes de ser aplicada.

Em síntese, este estudo fornece uma contribuição valiosa para a compreensão dos impactos da pandemia da COVID-19, destacando a importância da abordagem declarativa, adaptável a contextos imprevisíveis e instáveis.

Os dados e scripts utilizados neste trabalho estão disponíveis publicamente no Github, através da URL: <https://github.com/natalia-sendra/Covid-Mining-Process-Latin-America>.

6 Referências

AALST, W. (2016). Data Science in Action. In: Process Mining. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-49851-4_1

AALST, W.M.P. (2022a). Process Mining: A 360 Degree Overview. In: van der Aalst, W.M.P., Carmona, J. (eds) Process Mining Handbook. Lecture Notes in Business Information Processing, vol 448. Springer, Cham. https://doi.org/10.1007/978-3-031-08848-3_1

AALST, W.M.P. (2022b). Foundations of Process Discovery. In: van der Aalst, W.M.P., Carmona, J. (eds) Process Mining Handbook. Lecture Notes in Business Information Processing, vol 448. Springer, Cham. https://doi.org/10.1007/978-3-031-08848-3_

ACEVEDO I, *et al*, Informality in the time of COVID-19 in Latin America: Implications and policy options. PLoS One. 2021 Dec 16;16(12):e0261277. doi: 10.1371/journal.pone.0261277. PMID: 34914780; PMCID: PMC8675676.

ANNELIES, Wilder-Smith; OSMAN, S. (2020). Public Health Emergencies of International Concern: A Historic Overview.. Journal of Travel Medicine, 27(8) doi: 10.1093/JTM/TAAA227

ARAGHI, S. N. *et al.*, Stable heuristic miner: Applying statistical stability to discover the common patient pathways from location event logs, Intelligent Systems with Applications, Volume 14, 2022, 200071, ISSN 2667-3053, <https://doi.org/10.1016/j.iswa.2022.200071>.

AWOTUNDE, J. B., *et al.*, 8 - Application of big data in COVID-19 epidemic, Editor(s): Utku Kose, Deepak Gupta, Victor Hugo C. de Albuquerque, Ashish Khanna, Data Science for COVID-19, Academic Press, 2022, Pages 141-165, ISBN 9780323907699, <https://doi.org/10.1016/B978-0-323-90769-9.00023-2>.

BACK, C. O. *et al.*, (2020). Imperative versus Declarative Process Mining: An Empirical Comparison. Department of Computer Science, University of Copenhagen.

BISSON, S. The power of process mining in Power Automate. InfoWorld, 27 jul. 2023. Disponível em: <https://www.infoworld.com/article/3703213/the-power-of-process-mining-in-power-automate.html>. Acesso em: 03/11/2023.

BOGARÍN, A.; *et al.*, (2018). Discovering learning processes using Inductive Miner: A case study with Learning Management Systems (LMSs). *Psicothema*, 30(3), 322-329. doi: 10.7334/psicothema2018.116

CABASINO, M.P., *et al.* (2013). Introduction to Petri Nets. In: Seatzu, C., Silva, M., van Schuppen, J. (eds) *Control of Discrete-Event Systems. Lecture Notes in Control and Information Sciences*, vol 433. Springer, London. https://doi.org/10.1007/978-1-4471-4276-8_10

DI CICCIO, C.; MONTALI, M. (2022). Declarative Process Specifications: Reasoning, Discovery, Monitoring. In: van der Aalst, W.M.P., Carmona, J. (eds) *Process Mining Handbook. Lecture Notes in Business Information Processing*, vol 448. Springer, Cham. https://doi.org/10.1007/978-3-031-08848-3_4

DONADELLO, I. *et al.* Declare4Py: A Python Library for Declarative Process Mining. The BPM 2022 Demos & Resources Forum. 20th Business Process Management Conference, Münster, Germany, September 11-16, 2022

DUMAS, M., *et al.* (2018). Essential Process Modeling. In: *Fundamentals of Business Process Management*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-56509-4_3

ERDOGAN, T. G. ; TARHAN, A., 2018, "Systematic Mapping of Process Mining Studies in Healthcare," in *IEEE Access*, vol. 6, pp. 24543-24567, 2018, doi: 10.1109/ACCESS.2018.2831244.

FUNDO MONETÁRIO INTERNACIONAL. (2020). *Perspectivas econômicas: As*

Américas. A persistência da pandemia obscurece a recuperação. <https://www.imf.org/-/media/Files/Publications/REO/WHD/2020/Oct/Portuguese/textp.ashx>

GARCIA, C. dos S. *et al.*, Process mining techniques and applications – A systematic mapping study, *Expert Systems with Applications*, Volume 133, 2019, Pages 260-295, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2019.05.003>.

GOOGLE. COVID-19: Relatórios de mobilidade da comunidade. Google. Disponível em: <https://www.ghsindex.org/>. Acesso em: 29 de outubro, 2023.

GHS INDEX. Global Health Security Index. GHS. Disponível em: <https://www.google.com/covid19/mobility/>. Acesso em: 29 de outubro, 2023.

GOMES, A. F. D.; *et al.*, "Comparative Analysis of Process Mining Tools" (2021). CAPSI 2021 Proceedings. 4. <https://aisel.aisnet.org/capsi2021/4>

INTERNATIONAL LABOUR ORGANIZATION. ILOSTAT. ILOSTAT. Disponível em: <https://ilostat.ilo.org/>. Acesso em: 29 de outubro, 2023.

JAZBIK, Lucas Seixas. Deviance mining of online processes with non-atomic events in the COVID-19 domain. 2022. 70 f. Dissertação (Mestrado em Engenharia de Produção) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2022.

KESICI, C. *et al.*, (2022). A Systematic Literature Review of Studies Comparing Process Mining Tools. *International Journal of Information Technology and Computer Science*. 14. 1-14. 10.5815/ijitcs.2022.05.01.

LU, Y. *et al.* (2021). A Novel Approach to Discover Switch Behaviours in Process Mining. In: Leemans, S., Leopold, H. (eds) *Process Mining Workshops. ICPM 2020. Lecture Notes in Business Information Processing*, vol 406. Springer, Cham. https://doi.org/10.1007/978-3-030-72693-5_5

MAGGI, F. M. *et al.* (2011). User-guided discovery of declarative process models. In N. Chawla, I. King, & A. Sperduti (Eds.), Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2011, Paris, France, April 11-15, 2011) (pp. 192-199). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/CIDM.2011.5949297>

MAGGI, F.M. (2013). Declarative Process Mining with the Declare Component of ProM. International Conference on Business Process Management.

MAGGI, F.M. (2019). Declarative Process Mining. In: Sakr, S., Zomaya, A.Y. (eds) Encyclopedia of Big Data Technologies. Springer, Cham. https://doi.org/10.1007/978-3-319-77525-8_92

MENEZES, M. Dois anos da declaração da pandemia de COVID-19. Fundação Oswaldo Cruz (Fiocruz), 11 mar. 2022. Disponível em: <https://www.ioc.fiocruz.br/noticias/dois-anos-da-declaracao-da-pandemia-de-covid-19>. Acesso em: 27/10/2023

MERTENS, S. *et al.*, Towards a decision-aware declarative process modeling language for knowledge-intensive processes, Expert Systems with Applications, Volume 87, 2017, Pages 316-334, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2017.06.024>.

MUNOZ-GAMA, J. *et al.*, Process mining for healthcare: Characteristics and challenges, Journal of Biomedical Informatics, Volume 127, 2022, 103994, ISSN 1532-0464, <https://doi.org/10.1016/j.jbi.2022.103994>

OUR WORLD IN DATA. Our World in Data. Our World in Data. Disponível em: <https://ourworldindata.org/>. Acesso em: 29 de outubro, 2023.

ORGANIZAÇÃO MUNDIAL DA SAÚDE - OMS. World Health Statistics 2022. Genebra: OMS, 2022. ISBN 978-92-4-005114-0

ORGANIZAÇÃO PAN AMERICANA DE SAÚDE. Health in the Americas 2022: Overview of the Region of the Americas in the Context of the COVID-19 Pandemic. PAHO, 27 set. 2022. Disponível em: <https://iris.paho.org/handle/10665.2/56472>.

Acesso em: 29 de out, 2023

PESIC, M., *et al.*, "DECLARE: Full Support for Loosely-Structured Processes," 11th IEEE International Enterprise Distributed Object Computing Conference (EDOC 2007), Annapolis, MD, USA, 2007, pp. 287-287, doi: 10.1109/EDOC.2007.14.

PESTEL, N. WHO Declares COVID-19 a Pandemic. *Acta Bio Medica Atenei Parmensis*, v. 91, n. 1, p. 157-60, 2020., doi:10.23750/abm.v91i1.9397

POROUHAN, P. *et al.*, "Process and deviation exploration through Alpha-algorithm and Heuristic miner techniques," 2014 Twelfth International Conference on ICT and Knowledge Engineering, Bangkok, Thailand, 2014, pp. 83-89, doi: 10.1109/ICTKE.2014.7001540.

PROVOST, F. e FAWCETT, T. "Data Science and its Relationship to Big Data and Data-Driven Decision Making," *Big Data*, vol. 1, no. 1, pp. 51–59, 2013, doi: 10.1089/big.2013.1508.

ROJAS, E. *et al.*, Process mining in healthcare: A literature review, *Journal of Biomedical Informatics*, Volume 61, 2016, Pages 224-236, ISSN 1532-0464, <https://doi.org/10.1016/j.jbi.2016.04.007>.

ROVANI, M. *et al.*, Declarative process mining in healthcare, *Expert Systems with Applications*, Volume 42, Issue 23, 2015, Pages 9236-9251, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2015.07.040>.

SHCHERBAKOV, M. *et al.*, (2014). Lean Data Science Research Life Cycle: A Concept for Data Analysis Software Development. *Communications in Computer and Information Science*. 466. 708-716. 10.1007/978-3-319-11854-3_61.

SCHONIG, S. *et al.*, (2016). Efficient and Customisable Declarative Process Mining with SQL. 10.1007/978-3-319-39696-5_18.

SLAATS, T. Declarative and Hybrid Process Discovery: Recent Advances and Open Challenges. *J Data Semant* 9, 3–20 (2020). <https://doi.org/10.1007/s13740-020-00112->

THE WORLD BANK. World Bank Open Data. The World Bank. Disponível em: <https://data.worldbank.org/>. Acesso em: 29 de outubro, 2023.

UNITED NATIONS. United Nations. United Nations. Disponível em: <https://www.un.org/en/>. Acesso em: 29 de outubro, 2023.

UNIVERSITY OF OXFORD. The Oxford Covid-19 Government Response Tracker. bsg. Disponível em: <https://www.bsg.ox.ac.uk/research/covid-19-government-response-tracker>. Acesso em: 29 de outubro, 2023.

VOGELGESANG, T. *et al.*, (2022). Celonis PQL: A Query Language for Process Mining. In: Polyvyanyy, A. (eds) Process Querying Methods. Springer, Cham. https://doi.org/10.1007/978-3-030-92875-9_13

WASAN, S. K. *et al.*, The impact of data mining techniques on medical diagnostics. Data Science Journal, v. 5, p. 119-126, 2006.

ANEXO 1 – Templates Declarative Process mining

Regras utilizadas no trabalho, obtidas em Schonig *et al.*, (2016).

- Template Response

```
SELECT
    'response',
    TaskA,
    TaskB,
    (CAST(COUNT(*) AS FLOAT)/CAST((SELECT COUNT(*) FROM Log WHERE
    Task LIKE TaskA) AS FLOAT)) AS Support,
    ((CAST(COUNT(*) AS FLOAT)/CAST((SELECT COUNT(*) FROM Log
    WHERE
    Task LIKE TaskA) AS FLOAT)) * (CAST((SELECT COUNT(*) FROM
    (SELECT Instance FROM Log WHERE Task LIKE TaskA GROUP BY
    Instance)t2) AS FLOAT)/CAST((SELECT COUNT(*) FROM (SELECT
    Instance FROM Log GROUP BY Instance)t) AS FLOAT))) AS
    Confidence
FROM
    Log a, (SELECT a.Task AS TaskA, b.Task AS TaskB FROM Log
    a, Log b WHERE a.Task != b.Task GROUP BY a.Task, b.Task) x
WHERE
    a.Task = x.TaskA EXISTS (SELECT * FROM Log b WHERE b.
    Task = x.TaskB AND b.Instance = a.Instance AND b.Time ] > a
    .Time ])
GROUP BY
    x.TaskA, x.TaskB
```

- Template Alternate Response

```
SELECT
    'alternateResponse',
    TaskA,
    TaskB,
    (CAST(COUNT(*) AS FLOAT)/CAST((SELECT COUNT(*) FROM Log WHERE
    Task LIKE TaskA) AS FLOAT)) AS Support,
    ((CAST(COUNT(*) AS FLOAT)/CAST((SELECT COUNT(*) FROM Log WHERE
    Task LIKE TaskA) AS FLOAT)) * (CAST((SELECT COUNT(*) FROM
    (SELECT Instance FROM Log WHERE Task LIKE TaskA GROUP BY
    Instance)t2) AS FLOAT)/CAST((SELECT COUNT(*) FROM (SELECT
    Instance FROM Log GROUP BY Instance)t) AS FLOAT))) AS
    Confidence
FROM
```

```

Log a , (SELECT a . Task AS TaskA , b . Task AS TaskB FROM Log
a , Log b WHERE a . Task != b . Task GROUP BY a . Task , b . Task ) x
WHERE
a . Task = x . TaskA AND EXISTS (SELECT * FROM Log b WHERE b
. Task = x . TaskB AND b . I n s t a n c e = a . I n s t a n c e AND b . Time > a
. Time )
AND NOT EXISTS(SELECT * FROM Log b , Log c WHERE c .
I n s t a n c e = a . I n s t a n c e AND c . Task = x . TaskA AND b .
I n s t a n c e = a . I n s t a n c e AND b . Task = x . TaskB AND c . Time
> a . Time AND c . Time < b . Time )
GROUP BY x . TaskA , x . TaskB

```

- Template Precedence

```

SELECT
' Precedence ' ,
TaskA ,
TaskB ,
(CAST(COUNT( *) AS FLOAT) /CAST( (SELECT COUNT( *) FROM Log
WHERE Task LIKE TaskB ) AS FLOAT) ) AS Support ,
( (CAST(COUNT( *) AS FLOAT) /CAST( (SELECT COUNT( *) FROM Log
WHERE Task LIKE TaskB ) AS FLOAT) ) * (CAST( (SELECT COUNT( *) FROM
(SELECT I n s t a n c e FROM Log WHERE Task LIKE TaskB GROUP BY I n s t a n
c e ) t 2 ) AS FLOAT) /CAST( (SELECT COUNT( *) FROM (SELECT I n s t a n c e
FROM Log GROUP BY I n s t a n c e ) t ) AS FLOAT) ) ) AS Confidence
FROM
Log a , (SELECT a . Task AS TaskA , b . Task AS TaskB FROM Log a , Log b WHERE
a . Task != b . Task GROUP BY a . Task , b . Task ) x 5 WHERE a . Task = x . TaskB
AND EXISTS (SELECT * FROM Log b WHERE b . Task = x . TaskA AND b . I n s t
a n c e = a . I n s t a n c e AND b . Time < a . Time )
GROUP BY
x . TaskA , x . TaskB

```

- Template Alternate Precedence

```

SELECT
' alternatePrecedence ' ,
TaskA ,
TaskB ,
(CAST(COUNT( *) AS FLOAT) /CAST( (SELECT COUNT( *) FROM Log WHERE
Task LIKE TaskB ) AS FLOAT) ) AS Support ,
( (CAST(COUNT( *) AS FLOAT) /CAST( (SELECT COUNT( *) FROM Log WHERE
Task LIKE TaskB ) AS FLOAT) ) * (CAST( (SELECT COUNT( *) FROM (SELECT
I n s t a n c e FROM Log WHERE Task LIKE TaskB GROUP BY I n s t a n c e ) t 2 )

```

```

AS FLOAT)/CAST( (SELECT COUNT(*) FROM (SELECT Instance FROM Log
GROUP BY Instance)t) AS FLOAT) ) AS Confidence
FROM
Log a, (SELECT a.Task AS TaskA, b.Task AS TaskB FROM Log a, Log b WHERE
a.Task != b.Task GROUP BY a.Task, b.Task) x WHERE a.Task = x.TaskB
AND EXISTS (SELECT * FROM Log b WHERE b.Task = x.TaskA AND b.Instance = a.Instance AND b.Time < a.Time) 6 AND NOT EXISTS(SELECT *
FROM Log b, Log c
WHERE c.Instance = a.Instance AND c.Task = x.TaskB AND b.Instance = a.Instance AND b.Task = x.TaskA AND c.Time < a.Time AND c.
Time > b.Time)
GROUP BY
x.TaskA, x.TaskB

```

- Template Responded existence

```

SELECT
'respondedExistence',
TaskA,
TaskB,
(CAST(COUNT(*) AS FLOAT)/CAST( (SELECT COUNT(*) FROM Log WHERE
Task LIKE TaskA) AS FLOAT) ) AS Support,
( (CAST(COUNT(*) AS FLOAT)/CAST( (SELECT COUNT(*) FROM Log
WHERE Task LIKE TaskA) AS FLOAT) ) * (CAST( (SELECT COUNT(*) FROM
(SELECT Instance FROM Log WHERE Task LIKE TaskA GROUP BY
Instance)t2) AS FLOAT)/CAST( (SELECT COUNT(*) FROM (SELECT
Instance FROM Log GROUP BY Instance)t) AS FLOAT) ) ) AS
Confidence
FROM
Log a, (SELECT a.Task AS TaskA, b.Task AS TaskB FROM Log
a, Log b WHERE a.Task != b.Task GROUP BY a.Task, b.Task) x
WHERE
a.Task = x.TaskB AND EXISTS (SELECT * FROM Log b WHERE b
.Task = x.TaskA AND b.Instance = a.Instance)
GROUP BY
x.TaskA, x.TaskB

```

- Template Not Sucession

```

SELECT
'notSucession',
TaskA,
TaskB,
(CAST(COUNT(*) AS FLOAT)/CAST( (SELECT COUNT(*) FROM Log WHERE
Task LIKE TaskA) AS FLOAT) ) AS Support,

```

```

((CAST(COUNT(*) AS FLOAT)/CAST((SELECT COUNT(*) FROM Log WHERE
Task LIKE TaskA ) AS FLOAT)) * (CAST((SELECT COUNT(*) FROM (SELECT
Instance FROM Log WHERE Task LIKE TaskA GROUP BY Instance)t2)
AS FLOAT)/CAST((SELECT COUNT(*) FROM (SELECT Instance FROM Log
GROUP BY Instance)t) AS FLOAT))) AS Confidence
FROM
Log a, (SELECT a.Task AS TaskA, b.Task AS TaskB FROM Log a, Log b WHERE
a.Task != b.Task GROUP BY a.Task, b.Task) x 5 WHERE a.Task = x.TaskB
AND a.Time < ALL (SELECT Time FROM Log b WHERE b.Task = x.TaskA
AND b.Instance = a.Instance) AND EXISTS (SELECT * FROM Log b
WHERE
b.Task = x.TaskA AND b.Instance = a.Instance) 7 AND a.Time >
ALL(SELECT Time FROM Log b WHERE b.Task = x.TaskB AND b.Instance
= a.Instance)
GROUP BY
x.TaskA, x.TaskB

```