



Aplicação da etapa de *Business Understanding* do processo CRISP-DM para a análise de informações de folha de pagamento

Jonatas Carvalho Domingos¹

Centro Universitário Academia, Juiz de Fora, MG

Evaldo de Oliveira da Silva²

Centro Universitário Academia, Juiz de Fora, MG

Linha de Pesquisa: Engenharia de Software

RESUMO

O CRISP-DM foi concebido no final de 1996 com objetivo de oferecer etapas para compreensão e preparação de dados para modelos de mineração de dados. O CRISP-DM abre caminho no entendimento de etapas para compreensão e preparação de dados para modelos de mineração de dados e compreensão para especialistas em tecnologia no atendimento das exigências dos clientes. Inicialmente, o CRISP-DM não foi construído de forma teórica ou acadêmica, porém seu método teve sucesso devido a praticidade e aplicação prática do mundo real de como as pessoas mineram e analisam dados e, além disso, atualmente, serve como base para um processo de ciência de dados. O CRISP-DM pode fornecer uma abordagem estruturada para compreensão e preparação de dados para diferentes projetos que necessitam que os dados estejam prontos para análise de dados. Neste sentido, este trabalho apresenta um estudo de caso com a aplicação do CRISP-DM para compreensão e preparação de dados de folha de pagamento, onde deverá comunicar duas bases de dados de gestão de pessoas de forma a criar modelos estatísticos para a utilização empresarial corporativos. Espera-se que este trabalho explicita de forma prática o CRISP-DM e como trabalhos futuros propõe a implementação de metodologias de reestruturação de dados para análise de dados de folha de pagamento.

Palavras-chave: Gestão de projetos. Modelo de dados. Metodologia. Preparação de

¹ Discente do Curso de Engenharia de Software do Centro Universitário Academia – UniAcademia. Endereço: Rua Olegário Maciel 1835/308. Celular: (32) 98707-0742. E-mail: jonatas.domingos@gmail.com

² Docente do Curso de Engenharia de Software do Centro Universitário Academia. Orientador.



dados. Análise de dados. Mineração de dados.

ABSTRACT

CRISP-DM was conceived in late 1996 with the goal of providing steps for understanding and preparing data for data mining models. CRISP-DM paves the way in understanding steps for understanding and preparing data for data mining models and understanding for technology experts in meeting customer requirements. Initially, CRISP-DM was not built theoretically or academically, but its method was successful due to the practicality and real-world practical application of how people mine and analyze data, and furthermore, it currently serves as the basis for a data science process. CRISP-DM can provide a structured approach to understanding and preparing data for different projects that require data to be ready for data analysis. In this sense, this paper presents a case study with the application of CRISP-DM for understanding and preparation of payroll data, where it should communicate two people management databases in order to create statistical models for corporate business use. It is expected that this work will explain in a practical way the CRISP-DM and as future work proposes the implementation of data restructuring methodologies for payroll data analysis.

1 INTRODUÇÃO

Um projeto é um conjunto de atividades organizadas e coordenadas, com objetivo de alcançar um determinado resultado dentro de um prazo e com recursos estabelecidos. Projetos podem ser de diferentes tipos e tamanhos, e a gestão bem-feita deles é fundamental para garantir que os objetivos sejam alcançados dentro do prazo e com eficiência.

De acordo com Vargas (2009), um projeto pode ser subdividido em fases ou etapas de desenvolvimento. A compreensão dessas etapas permite ao gestor, entre outros benefícios, melhorar o ciclo de vida do projeto e manter o controle dos recursos gastos para atingir os objetivos pré-estabelecidos. O bom gerenciamento do ciclo de vida possibilita que seja avaliada uma série de situações previstas e não previstas que poderão acontecer durante as fases do projeto, independentemente de seu contexto ou área de atuação.

Com base no estudo de Wirth e Hipp (2000), a metodologia CRISP-DM³ (*Cross-Industry Standard Process for Data Mining*) é um processo que exige criatividade, habilidades e conhecimento, o qual não se apoia em uma ferramenta padronizada para a coleta de dados e informações.

³ CRISP-DM. Disponível em: <https://www.ibm.com/docs/it/SS3RA7_18.3.0/pdf/ModelerCRISPDM.pdf>. Acessado em: 01 fev. 2023.



De acordo com IBM (2021) (*International Business Machine*) em seu manual *IBM SPSS Modeler CRISP-DM – Guide*, a gestão de projeto e o CRISP-DM são duas metodologias distintas, porém complementares, utilizadas em projetos de ciência de dados. Considera-se que a utilização do CRISP-DM seja fundamentada em etapas (etapas iniciais) para condução de projetos que necessitem da compreensão e preparação de dados..

O CRISP-DM fornece uma estrutura clara para garantir que os dados sejam tratados de forma confiável e os modelos sejam construídos e avaliados adequadamente. Esta metodologia é composta por seis fases: *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation* e *Deployment*. Para a abordagem deste trabalho será utilizada a metodologia CRISP-DM, pois fornece uma abordagem estruturada para compreensão e preparação de dados (*Business Understanding*, *Data Understanding*, *Data Preparation*) para diferentes projetos que necessitam que os dados estejam prontos para diferentes análises.

Desta forma, este trabalho apresenta um estudo de caso com a aplicação do CRISP-DM para análise de indicadores de recursos humanos da empresa OCYAN.S.A.⁴, na qual o autor deste trabalho é funcionário da OCYAN que apresenta a necessidade de reformulação dos dados que estavam sendo integrados entre o seu sistema de gestão de pessoas com o sistema de folha de pagamento. Valores incorretos constavam nas tabelas dos sistemas gerando instabilidade na rotina administrativa dos setores de recursos humanos devido a excessivas inconsistências de informações dos funcionários relativas às suas ausências. Com o intuito de trabalhar o problema apresentado, aplica-se as etapas iniciais do CRISP-DM a fim de apresentar a reestruturação dos dados realizada sanando as divergências para melhoria do funcionamento dos sistemas.

Este trabalho apresenta como preparar registros em uma nova base de dados com uma modelagem diferente para a inserção das informações de ausências no sistema data master de gestão. A comunicação entre os sistemas se dá pelo uso de uma API (*Application Programming Interface*) de comunicação a qual consome e envia extrações no formato de tabelas transformadas em texto de bancos de dados disponibilizados na nuvem tanto do sistema de folha quanto do sistema de gestão. Será aplicado o método CRISP-DM para o entendimento entre as regras de negócios da empresa na representação dos dados relativos às informações corporativas, possibilitando a criação de métricas e análises estatísticas. Apresenta-se como resultado compreensão e preparação de dados de informações de folha de pagamento da

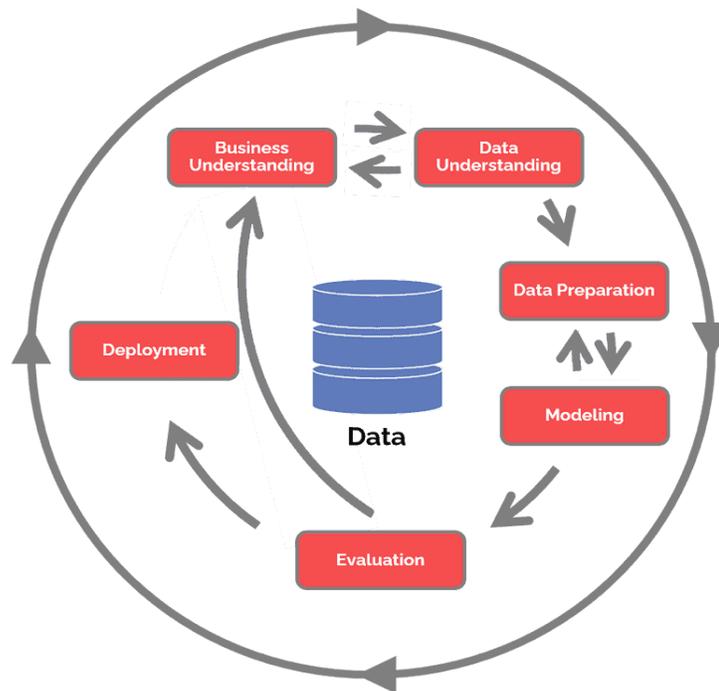
⁴ OCYAN.S.A. A Ocyan fornece soluções para a indústria de óleo e gás upstream offshore no Brasil e no exterior.

Ocyan.S.A; posteriormente e, por fim, as considerações finais e perspectivas de trabalhos futuros.

2 CRISP-DM

Esta seção apresenta as fases do CRISP-DM e como são conceituadas para diferentes projetos que necessitam da compreensão e preparação de dados. A Figura 1 representa, de uma forma geral, o processo proposto pelo CRISP-DM e suas etapas. Em seguida, as seções deste referencial teórico definem as etapas.

Figura 1. Representação do esquema que organiza as fases do CRISP-DM.



Fonte: HOTZ (2019), N. What is CRISP-DM?

2.1 . *Business Understanding*

A fase de *Business Understanding* (ou entendimento das regras do projeto) concentra-se na compreensão dos objetivos e requisitos do projeto. As tarefas nesta fase são atividades fundamentais e ocorrem na maioria dos projetos: Determinar os objetivos: É o entendimento completo do que realmente se deseja realizar, e, em seguida, definir os critérios de sucesso do projeto. Avaliar a situação: Determina a disponibilidade de recursos, os requisitos



do projeto, avaliação dos riscos e contingências, além do custo-benefício da realização do projeto. Determinar as metas de utilização dos dados, além de definir os objetivos, deve-se definir como é o sucesso de uma perspectiva técnica dos dados. Produzir o plano do projeto: Seleção das tecnologias e ferramentas e defina planos detalhados para cada fase do projeto.

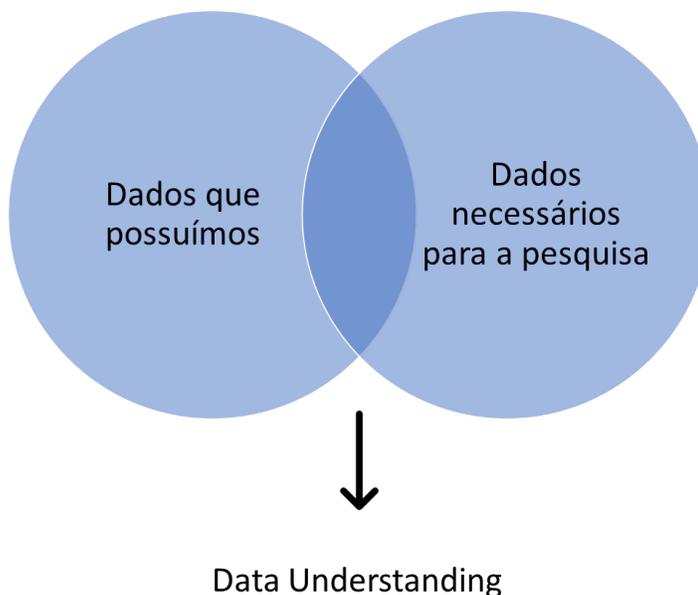
A necessidade deste projeto é o ajuste do controle de gestão dos dados de ausências dos funcionários da empresa Ocyan que informa aos departamentos de recursos humanos relatórios estatísticos, temporais e quantitativos da dinâmica do trabalho corporativo. Uma vez que o funcionamento do negócio se utiliza de escalas dinâmicas entre horas trabalhadas e embarques periódicos. Portanto informações estatísticas das ausências também são usadas como metas gerenciais para todo o controle das operações internas.

2.2 Data Understanding

Na fase de *Data Understanding* (Entendimento dos dados) é efetuado o relacionamento entre o dado da regra de negócio e as informações referentes à regra de negócio, o que de fato será utilizado, como está representado na Figura 2. Huber et.al (2019) estabelece o entendimento entre dados que possuímos em relação aos dados que precisamos para desenvolvimento da pesquisa. Algumas questões são feitas para obter o entendimento sobre os dados que necessitamos: Qual a disponibilidade dos dados? Qual a qualidade dos dados? Qual a granularidade? Ou seja, como aprofundar e detalhar os dados? Qual a frequência de atualização dos dados? Esta fase também direciona o foco para identificar, coletar e analisar os conjuntos de dados que podem ajudar no atingimento das metas do projeto. Esta fase também tem quatro tarefas:

- Coletar os dados iniciais: Aquisição dos dados necessários e (se necessário), utilizá-los em uma ferramenta de análise.
- Descrever os dados: Exame dos dados e documentação das suas propriedades, tais como: formato de dados, número de registros ou como os campos são identificados.
- Explorar os dados: Aprofundamento nos dados. Consulte-os, visualize-os e identifique relacionamentos entre os dados. Nesta tarefa hipóteses são formuladas.
- Verificar os dados: quão limpos/sujos estão os dados? Documentação de quaisquer problemas de qualidade. Identifica algumas questões envolvendo anomalias presentes nas coleções de dados.

Figura 2. Representação do conceito de Data Understanding.



Fonte: Elaboração própria.

Conforme Wirth e Hipp (2000), em um estudo experimental é de suma importância que seja compreendido o conceito de cada métrica a ser analisada e o que ela representa para a qualidade do *software*. Por isso, nesta seção serão abordados os conceitos de cada métrica utilizada neste estudo.

2.3. Data Preparation

A fase de Data Preparation (preparação dos dados), muitas vezes referida como manipulação de dados de acordo com Redman (2008), é um conceito abstrato pois não é possível tocar ou ver os dados, e o que manipulamos são registros de dados que vêm em um número quase ilimitado de formas, por exemplo: papel, tabelas impressas ou digitais e gráficos.



Na fase de preparação dos dados, não há uma técnica pré-definida pelo CRISP-DM, então podem ser usadas diferentes técnicas, umas delas é o ETL⁵ (*Extract Transform Load*), a qual foi escolhida neste trabalho. Savitri e Laksmiwati (2011) explicam que o ETL é executado para preparar o(s) conjunto(s) de dados final(is) para modelagem e é composto em cinco tarefas:

- Selecionar dados: Determinar quais conjuntos de dados serão usados e a documentação dos motivos da inclusão/exclusão.
- Estabelecer os dados limpos: É considerada a tarefa que pode levar mais tempo na preparação. Esta fase evita “lixos” que não podem ser utilizados nas ferramentas de análise, sob o risco de enviesar os dados analisados. Uma prática comum durante esta tarefa é corrigir, imputar ou remover valores incorretos.
- Construir dados: É derivação de novos atributos que serão úteis. Por exemplo, converter a matrícula e atributos chaves para identificação única.
- Integrar dados: Criação de novos conjuntos de dados combinando dados de várias fontes.
- Formatar dados: Os dados podem ser “reformatados”, caso necessário. Por exemplo, a conversão de valores de *string* que armazenam números em valores numéricos para poder realizar operações matemáticas.

2.4. Modeling

Na fase de modelagem são construídos e avaliados modelos com base em várias técnicas diferentes. Huber et.al (2019) aborda que as hipóteses levantadas na fase de *data understanding* são recuperadas, com o objetivo de utilizar o melhor modelo para solução dos problemas. São utilizados campos que servem para conduzir os resultados na execução do modelo.

Na criação do design, enquanto aguarda sua abordagem de modelagem, pode ser necessário dividir os dados em conjuntos de execução, teste e validação. Estabelecer critérios de avaliação pela existência de mais de um tipo de abordagem e técnicas que competem entre si, interpretando os resultados do modelo com base no conhecimento do domínio, nos critérios de sucesso predefinidos e no design do teste. O trabalho desta análise realizou as diferentes análises de dados. Por exemplo, as seguintes questões foram de interesse: Quais são as

⁵ ETL. Disponível em: < ieeexplore.ieee.org/document/6457795>. Acesso em 10 de jan 2023.



influências dos valores dos dados? Qual é a faixa ideal de execução do processo de integração dos dados? Como o processo afeta o dia a dia dos sistemas satélites?

2.5. *Evaluation e Deployment*

A fase de Avaliação analisa mais amplamente qual modelo atende melhor ao propósito do projeto e o que deve ser feito em seguida. Esta fase tem três tarefas:

- Avaliar os resultados: Os modelos atendem aos critérios de sucesso do projeto? Qual ou quais devemos aprovar?
- Revisar o processo: Revise o trabalho realizado. Alguma coisa foi esquecida? Todas as etapas foram executadas corretamente? Resuma as descobertas e corrija qualquer coisa, se necessário. Determinar as próximas etapas: Com base nas três etapas anteriores, determine se deve continuar com a implantação, iterar mais ou iniciar novos projetos. Um modelo não é particularmente útil a menos que o usuário ou especialista do domínio possa acessar seus resultados. A complexidade desta fase varia muito.
- A tarefa de Implantação (fase final) tem quatro fases:

Planejar a implantação: Desenvolva e documente um plano para implantar o modelo.
Planejar o monitoramento e manutenção: Desenvolva um plano completo de monitoramento e manutenção para evitar problemas durante a fase operacional (ou fase pós-projeto) de um modelo.
Produzir relatório final: A equipe do projeto documenta um resumo do projeto que pode incluir uma apresentação final dos resultados da mineração de dados.
Revisar do projeto: Conduza uma retrospectiva do projeto sobre o que deu certo, o que poderia ter sido melhor e como melhorar no futuro.

3. APLICAÇÃO DA ETAPA DE *BUSINESS UNDERSTANDING* DO PROCESSO CRISP-DM PARA ANÁLISE DE INFORMAÇÕES DE FOLHA DE PAGAMENTO

A realização deste trabalho de conclusão de curso fez parte de um projeto para compreensão e preparação de dados para geração de indicadores de recursos humanos. Para execução do projeto foram levantados prazos e fases para a condução de cada etapa do planejamento, embasado nas possibilidades orçamentárias previamente definidas pela companhia. Para os avanços das etapas deste projeto foi fundamental a otimização das análises das regras de negócio na fase de *Business Understanding*, que na prática foram reuniões de passagem de conhecimento pelos *stakeholders*, onde pela metodologia



CRISP-DM, é o momento que pode haver retrabalho caso a consequência de um baixo nível de análise do negócio ocorra e avance para as etapas de estruturação dos dados e modelagem. Houve uma grande utilização de tempo do projeto na etapa de análise do negócio, uma etapa que durou aproximadamente dois meses de constantes reuniões com os *stakeholders* para todo o levantamento necessário, foi considerado a fase de maior impacto para toda a vida deste projeto.

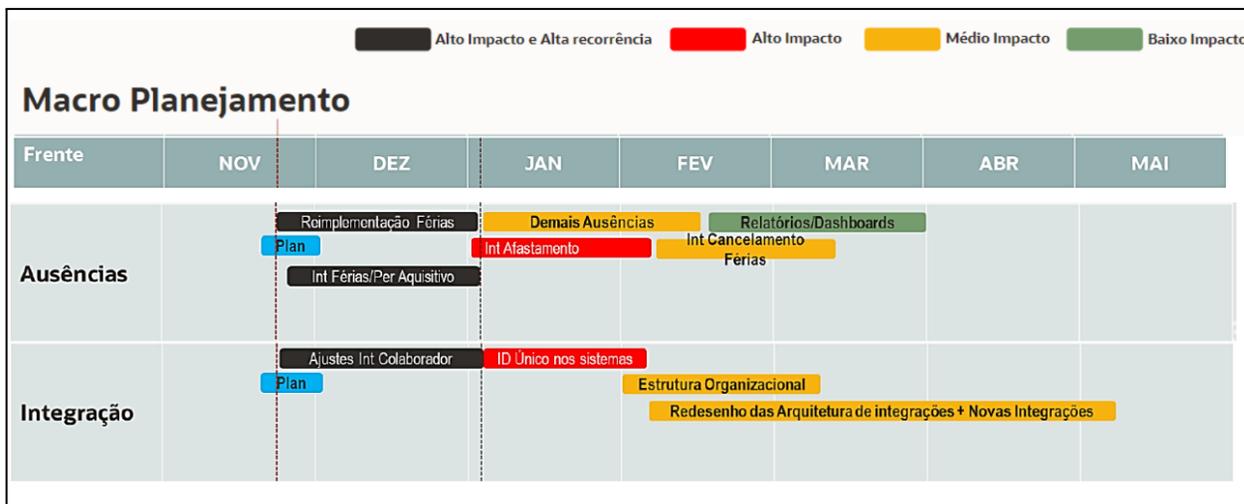
O projeto foi desenvolvido em uma plataforma *cloud* utilizada pela Ocyan no seu sistema de gestão de pessoas: *Oracle de Human Capital Management*⁶ (HCM), um produto de gestão empresarial no âmbito corporativo. Foi necessário pensar nas partes interessadas, *stakeholders* de toda operação onde os impactos seriam positivos ou negativos, cujo objetivo desta operação tinha como alvo. A análise de *stakeholders* dividiu duas linhas de frentes: *Stakeholders* primários e secundários (MARTINS, 2002), onde a ação dos primários foi em definir os recursos humanos necessários do projeto e aprovações das propostas implementadas. Já os secundários tinham como proposta o enfrentamento dos recursos implantados para uma melhoria proativa em tempo de projeto.

No macroplanejamento apresentado na Figura 3 foi considerado um trabalho de implementação de sete meses para a entrega, em duas frentes distintas, a implementação da gestão das ausências dos funcionários para o controle de gestão de pessoas e implantação das integrações das ausências para o sistema folha de pagamento. Foi utilizado todo o prazo estipulado para este projeto, considerando todas as fases e etapas de acompanhamento das entregas dos itens desenvolvidos para as equipes de suporte e *stakeholders* envolvidos no processo.

⁶ HCM. Disponível em: <<https://www.oracle.com/br/human-capital-management>>. Acessado em 10 de janeiro 2023.



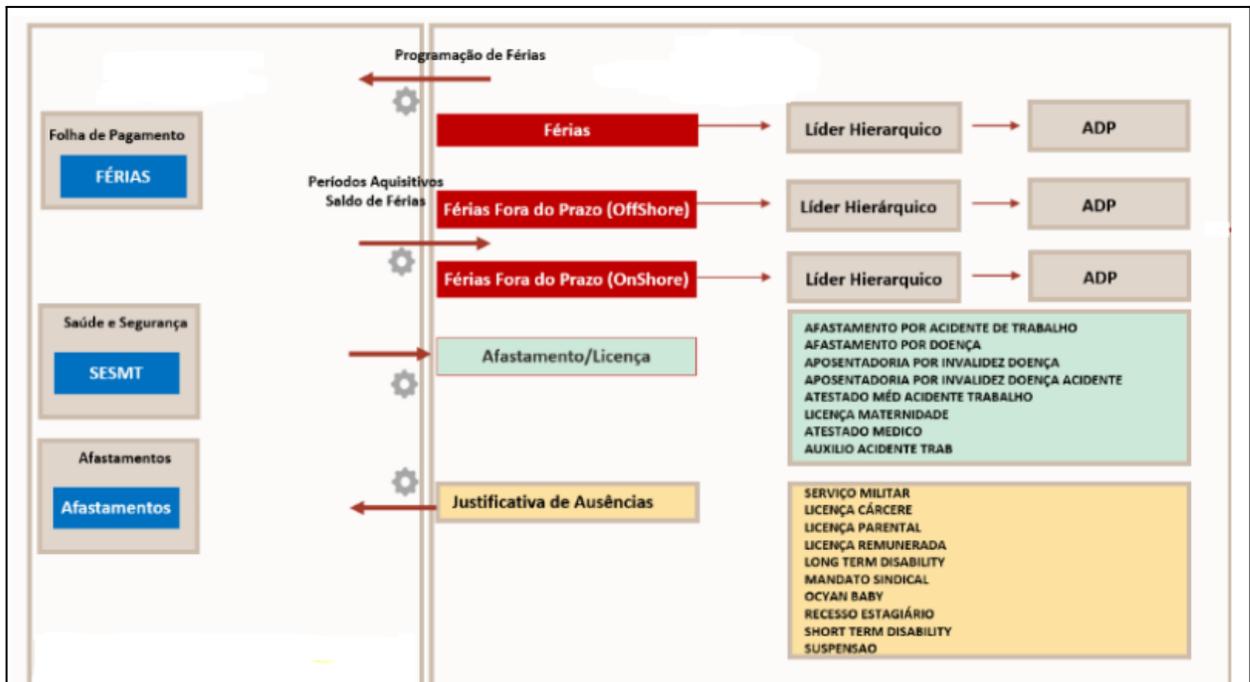
Figura 3. Macroplanejamento do projeto



Fonte: Elaboração própria

Os dados foram categorizados de acordo com as necessidades do negócio apresentadas pelos *stakeholders*, evitando conflito de regras entre setores da empresa. Toda uma base de dados corporativa foi utilizada, mantendo-se os históricos de funcionários ativos da empresa, respeitando o vínculo de cada funcionário. A Figura 4 detalha o sentido da integração dos dados e os fluxos das regras de negócios empresarial que envolvem o projeto. A integração incluiu características de unificação, consolidação e ações integradoras que foram importantes para a sincronização do processo no dia a dia do negócio, gerando um controle quantitativo dos erros de inconsistências entre a base do sistema de gestão de pessoas com o sistema de folha, nas fases de testes e na visão do *Business Understanding*

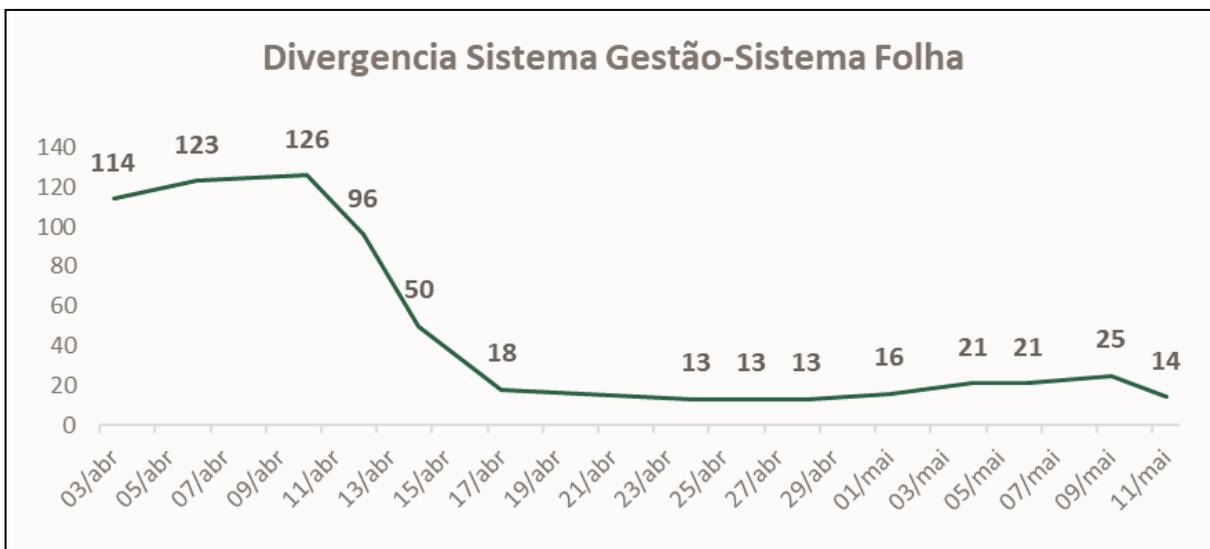
Figura 4. Representação das integrações de ausências



Fonte: Elaboração própria

A análise de erro de integração, conforme exemplificado na Figura 5, foi uma etapa crucial no desenvolvimento deste projeto, pois excessivas falhas podem desenvolver insatisfação e alta resistência dos usuários. Esta etapa consistiu em identificar e mapear as falhas na integração das diferentes partes do sistema respectivas às funções administrativas de cada usuário. Um exemplo: um simples *timing* de execução do processo de integração e a frequência das rotinas já poderia ocasionar uma falha para a equipe de recursos humanos. A partir do quinto mês de implantação do projeto foi verificada uma queda expressiva nas divergências entre o sistema de folha e o sistema de gestão de pessoas.

Figura 5. Representação das divergências de integrações de ausências



Fonte: Elaboração própria

Na fase dos testes foram executados 34 cenários de testes, conforme apresentado na Figura 6, com cada representante do negócio em específico quando se tratava de tarefas administrativas, estressando ao máximo as tarefas tanto dentro das regras propostas no plano de execução, quanto fora dos limites da regra de negócio solicitada.

Figura 6. Representação dos Testes



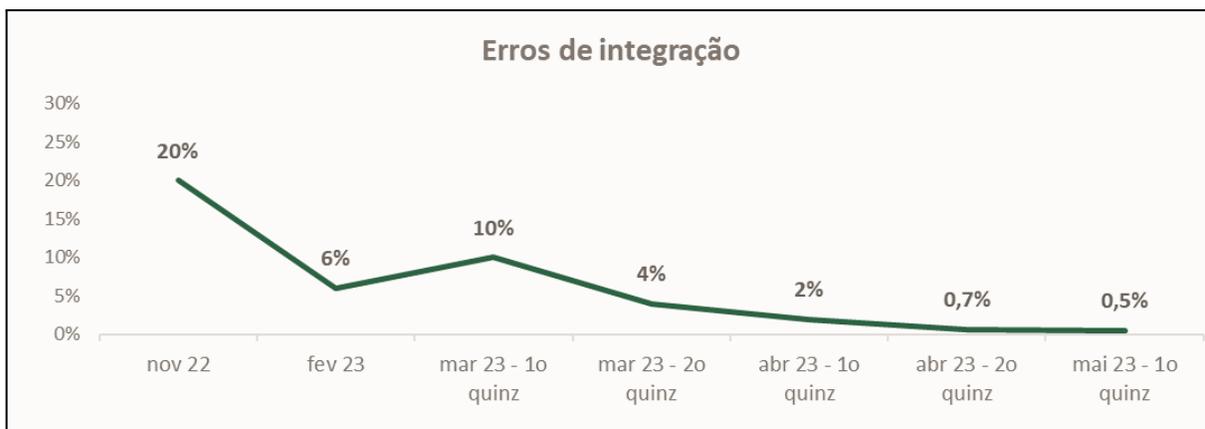
Fonte: Elaboração própria

Na fase de *hypercare* foi analisada uma diminuição considerável dos erros de integração baixando um volume de mais de 20%, considerando a equidade entre as bases sistêmicas e as



movimentações entre os dados. Observa-se na Figura 7, uma redução expressiva nos erros de integração, o que se deve ao entendimento pleno do negócio e suas regras corporativas.

Figura 7. Representação dos erros de integrações



Fonte: Elaboração própria

4. PREPARAÇÃO DE DADOS PARA ANÁLISE DE INFORMAÇÕES DE FOLHA DE PAGAMENTO DE DADOS COM MÉTODO USANDO CRISP-DM

Após a compreensão dos dados, o setor de Recursos Humanos da Ocyan apresentou divergências de integração de dados. Assim, houve a necessidade de preparar os dados para evitar problemas legais na geração de informações na folha de pagamento dos funcionários. Além disso, existe a necessidade de fazer com que os dados legais sejam integrados do sistema principal de gestão de pessoas para o sistema de folha de pagamento. As metas são manter a equidade das informações, manter ativo o ciclo de atualização entre os sistemas integradores e gerar processos e procedimentos para a regra de negócio empresarial na manipulação sistêmica mantendo funcional a atividade empresarial. Os dados compreendidos e preparados, visam melhorar os indicadores da área de recursos humanos.

No exemplo abaixo, a Tabela 1 apresenta registros de ausências pela base de dados do sistema de folha. Nesta etapa faremos a preparação dos dados para a comunicação entre os sistemas de folha e sistema de pessoas. Foram selecionados os atributos necessários pelo sistema de folha em comparação aos atributos chaves do sistema de gestão de pessoas, determinando quais conjuntos de dados seriam usados estabelecendo uma limpeza da estrutura de informações. Ao final, a equivalência da informação entre as tabelas foi construída



convertendo a matrícula e atributos chaves para identificação única, formatando os dados quando necessário. Os registros dos usuários possuem campo numérico no atributo matrícula, mas deveria estar constando um campo texto *string*. Já as informações de tempo, registro com o atributo com tipo data, não possui formatos compatíveis entre os sistemas. Neste caso, exige-se um esforço para trabalhar as informações, preencher os campos vazios e reformatar outros campos.

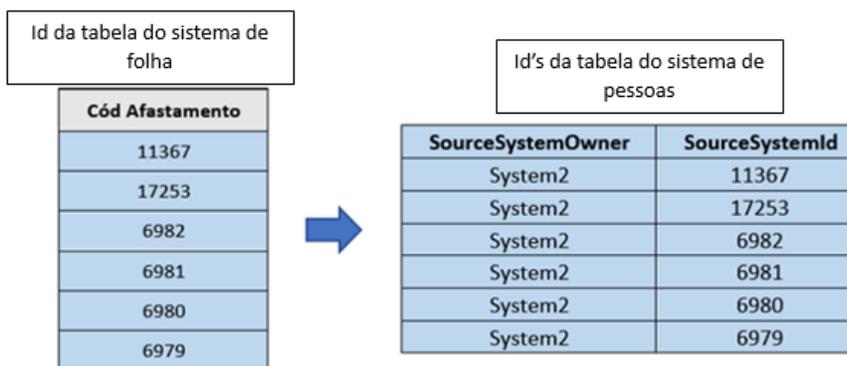
Tabela 1. Registro de ausência na folha de pagamento.

Cód. Afastamento	Data Inicial	Data Final	Trabalhador	PersonNumber	Matrícula	Afastamento	Ausência
11367	13/05/2022	15/05/2022	Nome 1	99999	999991	4	586 - ATESTADO MEDICO
17253	27/02/2023	03/03/2023	Nome 2	88888	888881	586	586 - ATESTADO MEDICO
6982	06/10/2020	28/02/2021	Nome 3	77777	777771	8	-
6981	25/07/2021	08/08/2021	Nome 3	77777	777772	3	-
6980	10/07/2020	23/07/2020	Nome 4	66666	666661	3	-
6979	11/04/2020	24/04/2020	Nome 4	66666	666662	3	-

Fonte: Elaboração própria.

A Figura 8 abaixo apresenta as diferenças entre a chave primária do registro do sistema de folha e a chave composta do sistema de gestão de pessoas, sendo necessário criar uma referência à fonte da informação para a futura comunicação entre os dados dos funcionários.

Figura 8. Representação do data Preparation do código do sistema de folha para o sistema de pessoas.



Fonte: Elaboração própria.



A Figura 9 abaixo apresenta o uso do campo de tipo data do sistema de folha e do sistema da gestão de pessoas. Os sistemas possuem considerável diferença quando se trata da manipulação deste campo, necessitando uma conversão do formato dd/mm/aaaa para aaaa/mm/dd e hh:mm permitindo a transferência e comunicação entre os sistemas.

Figura 9. Representação das datas entre os sistemas.

Atributo tipo data sistema de folha		Atributo tipo data do sistema de pessoas			
Data Inicial	Data Final	StartDate	StartTime	EndDate	EndTime
13/05/2022	15/05/2022	2022/05/13	00:00	2022/05/15	23:59
27/02/2023	03/03/2023	2023/02/27	00:00	2023/03/03	23:59
06/10/2010	28/02/2011	2010/10/06	00:00	2011/02/28	23:59
25/07/2021	08/08/2021	2021/07/25	00:00	2021/08/08	23:59
10/07/2020	23/07/2020	2020/07/10	00:00	2020/07/23	23:59
11/04/2020	24/04/2020	2020/04/11	00:00	2020/04/24	23:59



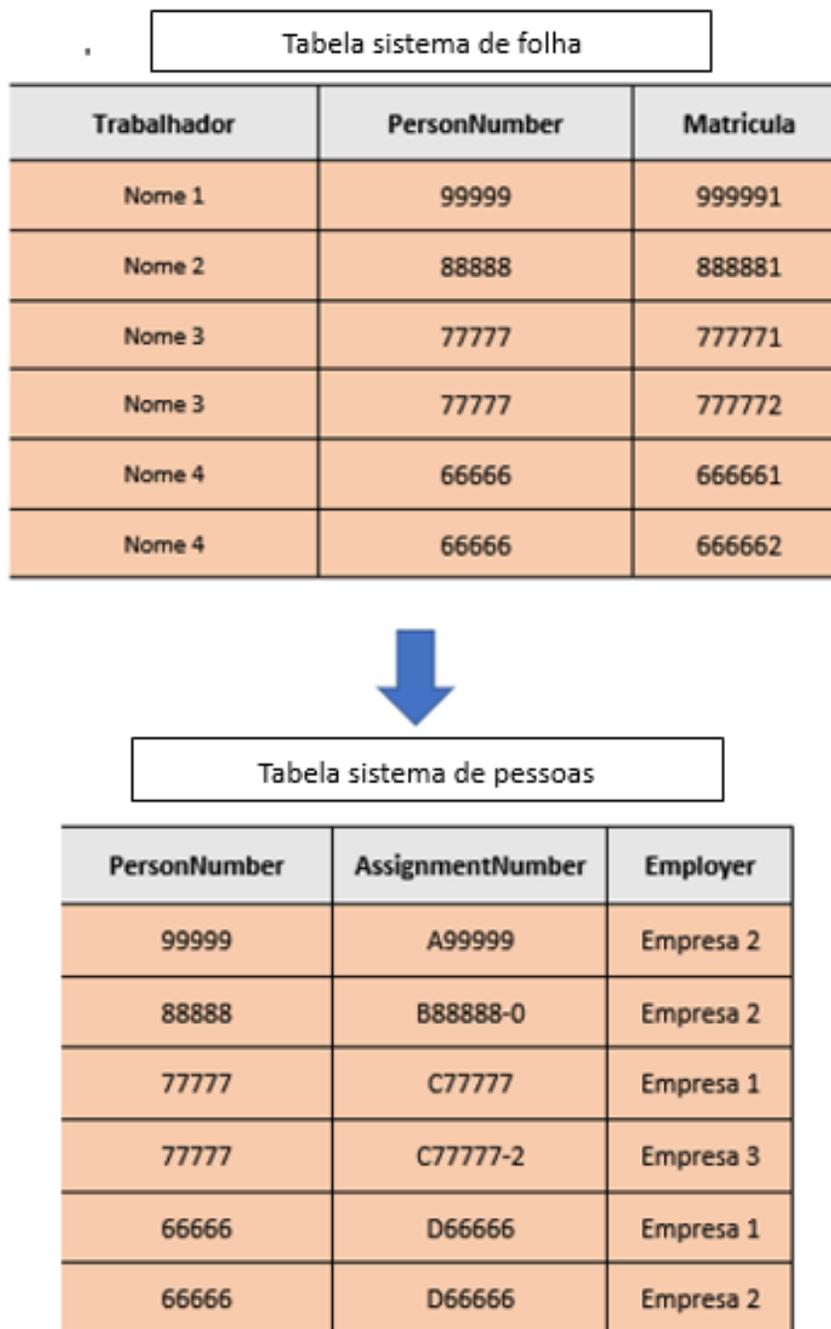
Fonte: Elaboração própria.

A figura 10 abaixo apresenta o uso dos atributos chaves de matrícula dos funcionários para os registros de ausências. O sistema de folha mantém somente o registro relativo à matrícula do funcionário e o sistema de gestão de pessoas possui os registros de matrícula e a empresa de origem, ou seja, qual foi a empresa que gerou as ausências.

A Figura 11 apresenta as diferenças entre a quantidade dos atributos relativos ao tipo da ausência dos funcionários, por seus códigos e status. O sistema de folha e o sistema de gestão de pessoas possuem, novamente, considerável diferença na tratativa do mesmo dado.



Figura 10. Representação da matrícula do funcionário entre os sistemas de folha e sistema de pessoas



Fonte: Elaboração própria.



Figura 11. Representação do tipo de ausência entre os sistemas

Tabela sistema de folha	
Afastamento	Ausência
4	586 - ATESTADO MEDICO
586	586 - ATESTADO MEDICO
8	-
3	-
3	-
3	-

↓

Tabela sistema de pessoas					
AbsenceType	Comments	ApprovalStatus	AbsenceStatus	<u>StartDate</u> <u>Duration</u>	<u>EndDate</u> <u>Duration</u>
Afastamento por Doença < 15	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1
Afastamento por Doença < 15	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1
Licença Não Remunerada	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1
Afastamento por Doença	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1
Afastamento por Doença	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1
Afastamento por Doença	Está na plataforma por evento de carga histórica	APPROVED	SUBMITTED	1	1

Fonte: Elaboração própria.

Os dados compreendidos, transformados e formatados foram de suma importância para o avanço de todo o restante deste projeto. Os relacionamentos sistêmicos e a manipulação sistêmica dos usuários não foram afetados devida a exata conversão das informações, pelo uso da etapa de *Data Preparation*. Contribuindo para todo o avanço e entrega dos prazos.



5. CONSIDERAÇÕES FINAIS

O presente trabalho de conclusão de curso buscou mostrar a aplicação das fases iniciais da metodologia CRISP-DM para execução de um projeto de tecnologia da informação da área de recursos humanos da empresa Ocyan S.A.. As etapas exigidas do método CRISP-DM ampliam a visão do negócio e consideram uma transição clara nos dados e informações utilizadas.

Na etapa de *Business Understanding* foi possível compreender, a partir dos stakeholders, as regras de negócio mais voltadas para o comportamento das ausências. Na fase de *Data Preparation* procurou-se ter um entendimento da estrutura dos dados e um planejamento de como trabalhá-los de forma a responder às regras de negócio compreendidas. As duas etapas juntas definiram o modelo correto de integração a ser implementado para sanar as divergências apresentadas.

Devido a este projeto foram criadas melhorias na identificação entre matrículas e atributos de identificação de outras demandas empresariais, por conta da limpeza e transformação das informações na fase de preparação de dados. Em demandas futuras semelhantes, será aplicada novamente a metodologia CRISP-DM para execução de novos projetos de sistemas de TI, pois foi compreendida pelo negócio da Ocyan como uma abordagem completa para a gestão e manipulação da informação entre sistemas.

No presente momento pode ser considerado como totalmente sanadas as inconsistências de informações que uma vez ocorreram, que dificultavam o trabalho das métricas estatísticas operacionais para o cumprimento de avaliações gerenciais das informações de ausências dos funcionários. Portanto, para demandas análogas com complexidade semelhante em empresas ou indústrias, as análises e etapas do método CRISP_DM, se bem aplicadas, podem trazer bons resultados na estruturação das informações de negócio.

REFERÊNCIAS

HOTZ, Nick. What is CRISP DM? **Data Science Process Alliance**, 2019. Disponível em: <<https://www.datascience-pm.com/crisp-dm-2>>. Acesso em: 10 de dez. de 2022.

HUBER, Steffen. *et al.* DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. **Procedia CIRP**, n. 79, 403–408, 2019.



IBM (2021). **IBM SPSS Modeler CRISP-DM Guide**. Disponível em: <https://www.ibm.com/docs/it/SS3RA7_18.3.0/pdf/ModelerCRISPDM.pdf>. Acesso em: 01 fev. 2023.

MARTINS, J.C.C. **Gestão de projetos de desenvolvimento de software: PMI-UML**. Brasport, 2002.

REDMAN, Thomas C. (2008). **Data Driven: Creating a Data Culture**. Harvard Business Press.

SAVITRI, Fivien. N.; LAKSMIWATI, Hira. **Study of localized data cleansing process for ETL performance improvement in independent datamart**. IEEE, 2011.

VARGAS, Ricardo. (2009). **Gerenciamento de Projetos: Estabelecendo Diferenciais Competitivos**, 7^a edição, Rio de Janeiro, Brasport, 2009.

WIRTH, Rüdiger; HIPPE, Jochen. **CRISP-DM: Towards a standard process model for data mining**. Research Gate, 2000.