

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

FELIPE FRIGHETTO

**Projeto de desenvolvimento de um
dashboard de gestão utilizando data
analytics lifecycle**

Monografia de Conclusão de Curso apresentada
como requisito parcial para a obtenção do grau
de Especialista em Engenharia de Software e
Inovação

Orientador: Prof. Dra. Daniela Francisco Brauner

Porto Alegre
2021

CIP — CATALOGAÇÃO NA PUBLICAÇÃO

Frighetto, Felipe

Projeto de desenvolvimento de um dashboard de gestão utilizando data analytics lifecycle / Felipe Frighetto. – Porto Alegre: PPGC da UFRGS, 2021.

44 f.: il.

Monografia (especialização) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR-RS, 2021. Orientador: Daniela Francisco Brauner.

1. DADCanvas. 2.

Big Data. 3.

Data Science. 4.

Dashboard. 5.

Data Analytics Lifecycle. 6.

Microsoft Power BI. 7.

Data Analytics Lifecycle. I. Brauner, Daniela Francisco.
II. Título.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Rui Vicente Oppermann

Vice-Reitora: Prof^a. Jane Fraga Tutikian

Pró-Reitor de Pós-Graduação: Prof. Celso Giannetti Loureiro Chaves

Diretora do Instituto de Informática: Prof^a. Carla Maria Dal Sasso Freitas

Coordenador do PPGC: Prof. João Luiz Dihl Comba

Bibliotecária-chefe do Instituto de Informática: Beatriz Regina Bastos Haro

*“O importante é não parar de questionar.
A curiosidade tem sua própria razão de existir.”*

— ALBERT EINSTEIN

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a Deus, que permitiu ter saúde para enfrentar a pandemia do COVID-19, e fez com que meus objetivos fossem alcançados, durante todo o período de estudo.

Ao meu gestor e colegas de trabalho, que me incentivaram, me motivaram e me deram condições para a realização do trabalho de forma tranquila e com sabedoria.

Aos meus pais, que sempre me apoiam nos estudos, com sua experiência de vida, me ensinam muito, e com poucas palavras dizem o que é preciso escutar para seguir em frente.

A minha esposa, por compreender que por muitas vezes precisei ficar ausente para me dedicar aos estudos, mas que sempre me apoia em tudo que faço, pois sabe que é necessário para minha profissão e estabilidade da família.

Ao meu filho, que sempre diz “Papai você tem que ir estudar, não pode deixar para amanhã os trabalhos!”. Pequenas palavras, mas de extrema importância, pois apesar da idade, sabe discernir o certo e o errado e torna-se uma criança com muitos princípios.

A minha orientadora, Daniela F. Brauner, que não mediu esforços para me auxiliar na orientação deste trabalho, sem medo de errar, é uma profissional sem igual.

RESUMO

Com um volume excessivo de dados, sua variedade e sua velocidade de geração, surgem as necessidades de tomada de decisão mais rápidas e assertivas. A dificuldade em estabelecer o objetivo do problema do negócio e então aplicar métricas para saná-las, faz o *Data Science* se destacar ao propósito. Com o *Data Science* é possível realizar estudos de análises de problemas, e assim auxiliar as organizações a lidar nas suas demandas por informações para a tomada de decisão. Este trabalho propõe uma pesquisa de caráter exploratório, visando gerar conhecimento a partir do desenvolvimento de um painel de gestão para a área de planejamento e controle da produção para uma empresa metalúrgica, desenvolvido no *Microsoft Power BI*, aplicando o *Data Analytics Lifecycle*, e destacando na Fase 1 do DAL, a aplicação do DADCanvas, o que visa à compreensão dos objetivos de negócios.

Palavras-chave: DADCanvas. *Big Data*. *Data Science*. *Dashboard*. *Data Analytics Lifecycle*. *Microsoft Power BI*. *Data Analytics Lifecycle*.

ABSTRACT

With an excessive volume of data, its variety and its generation speed, the needs for faster and more assertive decision. The difficulty in establishing the objective of the business problem and then applying metrics to remedy it, makes Data Science stand out for its purpose. With Data Science it's possible to carry out problem analysis studies, and thus help organizations to deal with their demands for information for decision making. This work proposes an exploratory research, aiming to generate knowledge from the development of a management panel for the area of production planning and control for a metallurgical company, developed in Microsoft Power BI, applying Data Analytics Lifecycle, and highlighting in Phase 1 of the DAL, the application of DADCanvas, which aims to understand the business objectives.

Keywords: DADCanvas, Big Data, Data Science, Dashboard, Data Analytics Lifecycle, Microsoft Power BI, Data Analytics Lifecycle.

LISTA DE ABREVIATURAS E SIGLAS

ERP	<i>Enterprise Resource Planning</i>
BI	<i>Business Intelligence</i>
PCP	Planejamento e Controle da Produção
Infor LN	<i>Software ERP de Gestão da Empresa</i>
DOD	Tomada de decisão orientada por dados (DOD)
DAL	<i>Data Analytics Lifecycle</i>
DADCanvas	<i>Data Analytics Discovery</i>
GD	Grupo Diretivo
MRP	<i>Material Requirement Planning</i>
XML	<i>Extensible Markup Language</i>

LISTA DE FIGURAS

Figura 2.1	Categoria de Dados.....	15
Figura 2.2	Tipos de Dados	15
Figura 2.3	<i>Data Science</i> no contexto dos diversos processos relacionados a dados na organização	17
Figura 2.4	Ciclo de Vida <i>Data Analytics Lifecycle</i>	19
Figura 2.5	<i>Data Analytics Discovery</i> - DADCanvas.....	21
Figura 2.6	Comparativo <i>Business Intelligence</i> e <i>Data Science</i>	22
Figura 4.1	1º Versão do DADCanvas.....	28
Figura 4.2	2º Versão do DADCanvas.....	29
Figura 4.3	<i>Dashboard</i> de Gestão	32
Figura 4.4	Cabeçalho do Painel	32
Figura 4.5	Filtros do Painel.....	33
Figura 4.6	Indicadores de Grandezas	35
Figura 4.7	Gráfico Tempo Médio de Processamento de SC's e OC's.....	36
Figura 4.8	Gráfico Quantidade de Itens x Tipos de Problemas.....	37
Figura 4.9	Tabela Detalhes.....	39

LISTA DE TABELAS

Tabela 4.1	1º Versão 5 why's do problema	28
Tabela 4.2	2º Versão 5 why's do problema	30
Tabela 4.3	Exemplo de uma Estrutura de Engenharia de um produto manufaturado	38

SUMÁRIO

1 INTRODUÇÃO	11
1.1 Objetivos	12
1.2 Justificativa	13
2 REVISÃO TEÓRICA	14
2.1 <i>Big Data</i>	14
2.2 <i>Data Science</i>	16
2.3 Ciclo de Vida do <i>Data Science</i>	18
2.4 <i>Data Analytics Lifecycle</i>	18
2.5 DADCanvas	20
2.6 <i>Business Intelligence</i>	21
3 METODOLOGIA	24
3.1 Fase 1 - Descoberta (<i>Discovery</i>)	24
3.2 Fase 2 - Preparação dos Dados (<i>Data Preparation</i>)	24
3.3 Fase 3 - Planejamento do Modelo (<i>Model Preparation</i>)	25
3.4 Fase 4 - Construção do Modelo (<i>Model Building</i>)	25
3.5 Fase 5 - Comunicar os Resultado (<i>Communicate Results</i>)	25
3.6 Fase 6 - Operacionalizar (<i>Operationalize</i>)	25
4 APRESENTAÇÃO DO <i>DASHBOARD</i> DE GESTÃO	26
5 CONSIDERAÇÕES FINAIS	42
REFERÊNCIAS	44

1 INTRODUÇÃO

A disseminação do uso de tecnologias trouxe desafios em lidar com a grande quantidade de dados, colocando em evidência os conceitos de *Big Data* e *Data Science*. O *Data Science* está diretamente relacionada ao *Big Data*, pois surge justamente pelas características especiais relacionadas com o volume, variedade, velocidade e variabilidade dos dados (DHAR, 2013). Ela se destaca então ao ajudar as empresas a lidarem com essa nova sistemática de análise, gerenciamento e extração de informação a partir dos dados (CHANG; GRADY, 2019).

Segundo (CHANG; GRADY, 2019), o *Data Science* é a extração de conhecimento acionável diretamente dos dados por meio de um processo de descoberta ou formulação de hipótese e teste de hipótese, e portanto, é capaz de auxiliar as empresas a lidar com essa nova sistemática de análise dos dados, permitindo a extração de conhecimento a partir dos dados de forma mais eficaz. (MOURA, 2018).

Neste trabalho, serão aplicados os conceitos de *Data Science* para auxiliar a tomada de decisão de uma empresa metalúrgica. Vale salientar que no ano de 2012, com 75 anos de existência, e mais de 200 funcionários, a empresa atualizou seu antigo ERP personalizado, para um ERP de mercado (Infor LN), utilizando as melhores práticas de mercado, sem personalização. Após a implantação do sistema, muitos processos adotados se tornaram morosos e sem padrão, tornando assim, a tomada de decisão da empresa mais difícil, pois as formas de buscar informações no novo sistema não ficaram claramente definidas. No ano de 2015, os processos das áreas da empresa foram reestruturados com ênfase nas práticas e padrões do mercado. Após ações de melhorias de processos e definição de padrões, a base de dados empresarial ficou mais acessível, com padrões de processos, e dados confiáveis para a tomada de decisão. Ainda assim, a adoção de dados para apoiar a tomada de decisões são lentas, sem recursos visuais, e muitas vezes não são indicadas para resolução de problemas, acabando desconexa dos processos e objetivos de negócio.

Neste contexto, este trabalho utilizará os processos de tomada de decisão da área de PCP (Planejamento e Controle da Produção) da empresa como ponto de partida, pois ela é a área responsável pela maior parte das decisões do dia-a-dia, concentrando o gerenciamento dos projetos de fabricação, gerenciamento de montagens externas, programação de materiais, *follow-up* e gestão da informação, tudo para o alcance das exigências competitivas do mercado. É importante saber que, o PCP é uma função complexa que, na

concepção de (ERDMANN, 1998) deve ser decomposta em três etapas, abrangendo atividades de planejamento (longo prazo), programação (curto prazo antecedendo a produção) e controle (curto prazo, durante e após a produção). Com a concentração das diversas fontes de informações na área do PCP, é eminente a necessidade de investigar nos processos de *Data Science* formas para auxiliar a extração dos dados relevantes e transformá-los em conhecimento para a tomada de decisão na empresa.

Este trabalho visa desenvolver um projeto de *Data Science* para criação de um painel de controle de gestão *dashboard* para apoiar a tomada de decisão da área de planejamento e controle da produção na empresa estudada. Para conduzir um projeto de *Data Science* é necessário seguir um processo passo a passo para organizar as atividades e tarefas envolvidas na aquisição, processamento, análise e construção do produto de dados (MOURA, 2018). Para desenvolvimento do projeto, será utilizado o *Data Analytics Lifecycle*, desenvolvido para o curso *Data Science e Big Data Analytics da EMC*. (DIETRICH D.; HELLER, 2015).

1.1 Objetivos

Este trabalho tem por objetivo desenvolver um projeto de um *dashboard* de gestão para a área do PCP, utilizando o *Data Analytics Lifecycle* e o *Microsoft Power BI*. Os objetivos específicos, são:

- 1) Identificar questões de negócios importantes para a tomada de decisão na área de PCP, que possam ser respondidas com análise descritiva de dados;
- 2) Identificar bases de dados internas da empresa e externas, que possam auxiliar no processo de desenvolvimento do *dashboard*;
- 3) Identificar integrações de sistemas de informação necessárias para acesso às bases, desenvolvimento e implantação do *dashboard* na empresa;
- 4) Validar a aplicação *Data Analytics Lifecycle* como processo de desenvolvimento de um painel de dados para o PCP;
- 5) Definir e avaliar as hipóteses de análise de dados para tomada de decisão do PCP (se as perguntas identificadas geraram análises e foram respondidas a partir dos dados coletados).

1.2 Justificativa

No ano de 2012 uma empresa metalúrgica, com 75 anos de existência, e mais de 200 funcionários, atualizou seu ERP que anteriormente era 100 % customizado, para um ERP de mercado (Infor LN), com o único objetivo de customização zero.

Após a implantação do sistema, muitos processos adotados se tornaram morosos e sem padrão, tornando assim a tomada de decisão da empresa muito dificultosa, pois as buscas das informações no sistema eram desconhecidas e sem regras claras definidas.

No ano de 2015, os processos das áreas da empresa foram reestruturados com ênfase nas práticas e padrões do mercado. Após desenvolver diversos trabalhos de melhoria de processos e padrões, foi possível criar uma base de dados sólida, com padrões de processos, e dados de entradas e saídas confiáveis para a tomada de decisão, porém a geração destas análises para tomada de decisão são muito lentas, com nenhum recurso visual, e muitas vezes desconhecidas e desconexas aos objetivos de negócio.

Segundo (CHANG; GRADY, 2019) , *Data Science* é a extração de conhecimento acionável diretamente dos dados por meio de um processo de descoberta, ou formulação de hipóteses e teste de hipóteses, portanto este trabalho visa desenvolver um projeto de *Data Science* para criação de um painel de controle de gestão *dashboard* para apoiar a tomada de decisão da área do PCP na empresa estudada.

2 REVISÃO TEÓRICA

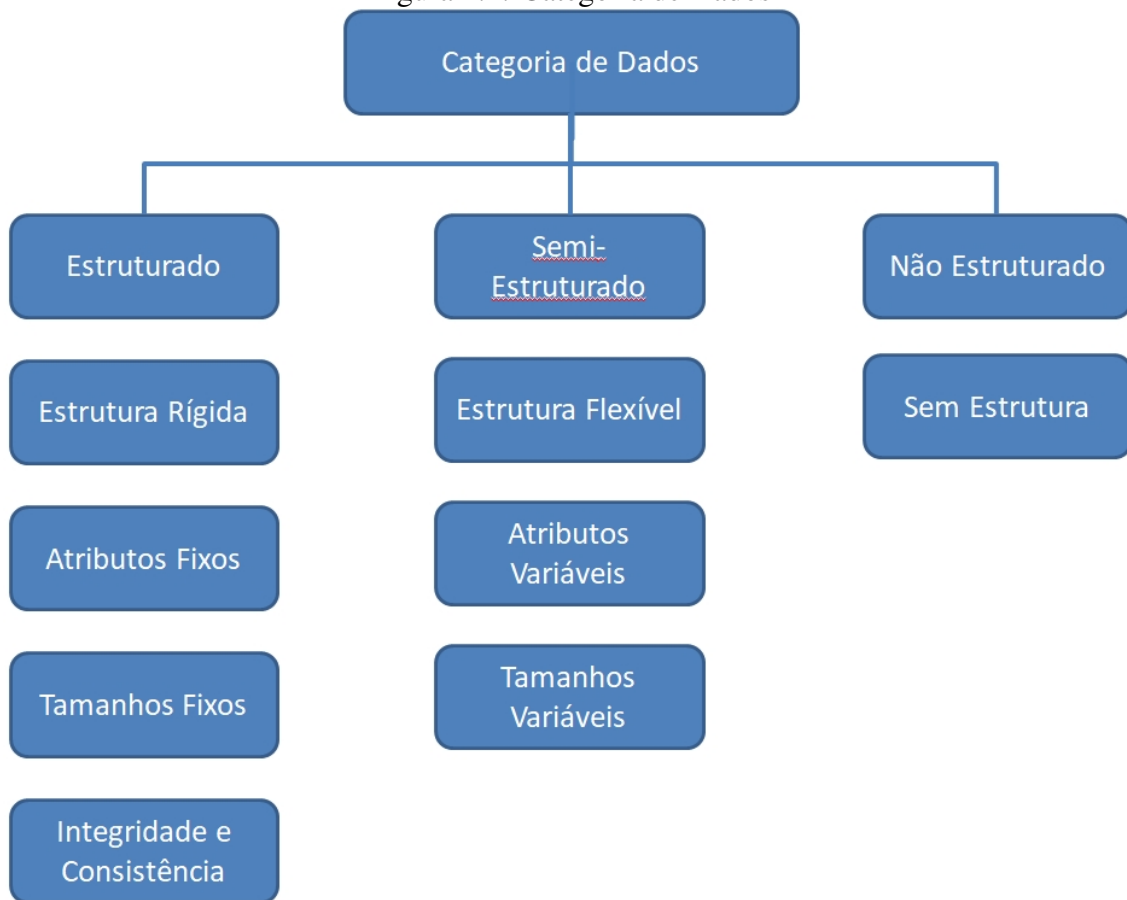
Os Dados, informação e conhecimento são elementos fundamentais para a comunicação e a tomada de decisão nas organizações(ANGELONI, 2003). A competitividade das empresas e a transformação dos dados em conhecimento, faz com que as empresas invistam tempo e dinheiro em *Data Science*, assim podem tomar melhores decisões no futuro. Para uma melhor compreensão dos valores potencializados dos dados, é muito importante saber sua origem, seus tratamentos, seu ciclo de vida, e porque muitas empresas estão investindo nesta área de *Data Science*. Este capítulo esta dividido em 6 seções: *Big Data*, *Data Science*, Ciclo de Vida da *Data Science*, *Data Analytics Lifecycle*, *DAD-Canvas* e *Business Intelligence*.

2.1 *Big Data*

O *Big Data* surgiu em 1997, devido ao grande aparecimento de dados, com uma maior variedade, volumes crescentes e velocidade de geração, dadas pelas mais diversas fontes geradoras de dados, como: mídias sociais, sensores de celulares, imagens médicas entre outras.

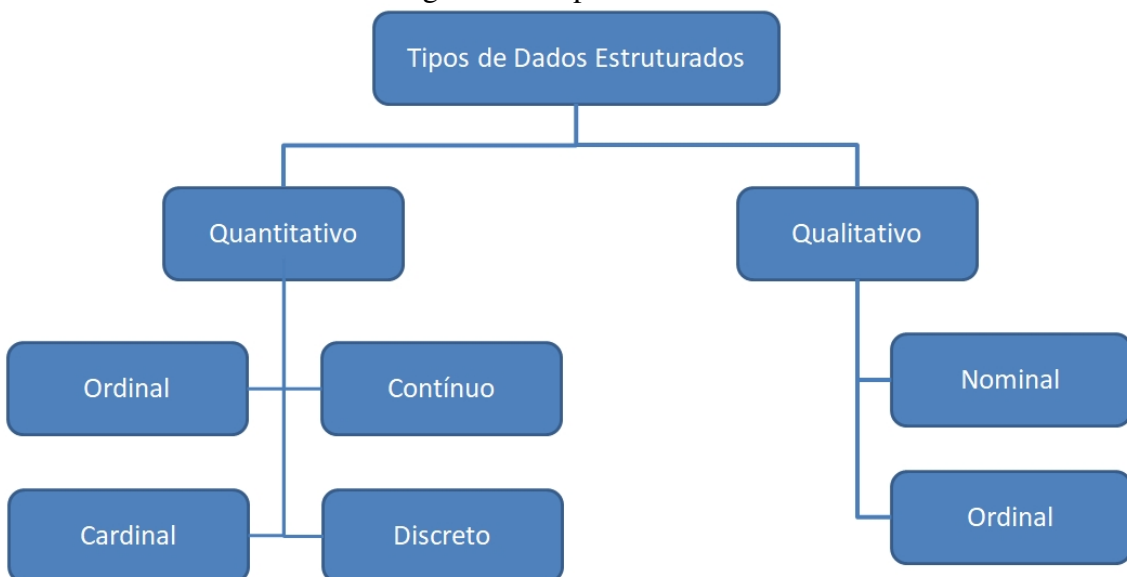
O elemento principal do *Big Data* são os dados. Segundo (ANGELONI, 2003) , dados são um conjunto de elementos ou ocorrências em estado bruto, ou seja, não possuem significado quando não vinculados com a realidade, é importante destacar que existem categorias de dados apresentado na figura 2.1, e tipos de dados apresentado na figura 2.2.

Figura 2.1: Categoria de Dados



Fonte: (SILVA, 2015) Adaptado pelo Autor

Figura 2.2: Tipos de Dados



Fonte: (SILVA, 2015) Adaptado pelo Autor

Com uma maior quantidade de dados gerados, e com uma maior dificuldade de extrair informações, foi necessário desenvolver melhores e mais rápidas ferramentas para lidar com os 5 V's, elementos importantes para o desenvolvimento do *Big Data*.

- Volume: é a característica mais relacionada ao *Big Data*, sinalizando o tamanho dos dados armazenados em recursos computacionais (CHANG; GRADY, 2019).
- Velocidade: é a taxa com que os dados são criados, armazenados, analisados e visualizados. Isso significa uma enorme quantidade de dados sendo processados em um curto período de tempo (CHANG; GRADY, 2019).
- Variedade: são os diferentes tipos de dados disponíveis para análise. Eles podem ser classificados como: estruturados (armazenados em banco de dados relacionais), não estruturados (texto livre) ou semiestruturados. (ALMEIDA. . . , 2002)
- Veracidade: esta diretamente ligada à fidelidade dos dados, ou quanto eles são verdadeiros;
- Valor: valor da informação gerada para a empresa, ou quanto a informação é útil.

2.2 *Data Science*

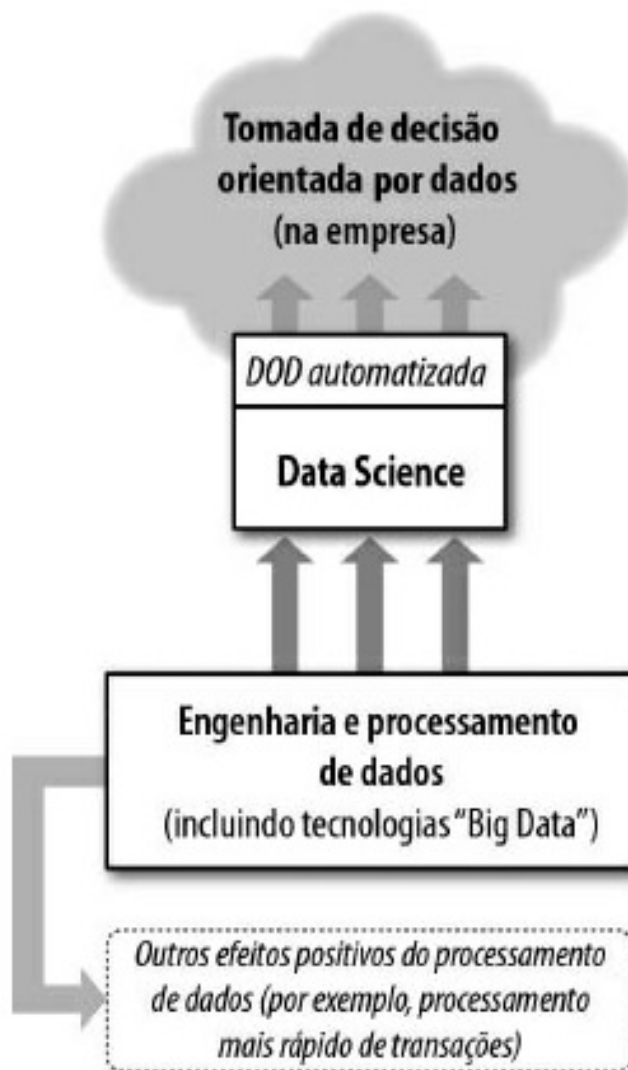
Com um crescente aumento nas quantidades de dados, surge à necessidade de analisar e extrair informações úteis, seja para os negócios, ou prever e descobrir o que vai ou pode acontecer no futuro.

O *Data Science* destaca-se nas organizações, tornando-se vital para a inteligência corporativa, pois assim conseguem utilizar a informações extraídas de seus dados, e aplicam melhores investimentos, ações economizando assim tempo e dinheiro.

Para que estas informações sejam extraídas com qualidade e eficiência, algumas técnicas e teorias das áreas da computação, engenharia, matemática, estatística, economia, mineração de dados, programação de computadores, inteligência artificial entre outras são aplicadas.

Segundo (PROVOST, 2016), o *Data Science* precisa ter acesso aos dados e, muitas vezes, beneficia-se da sofisticada engenharia de dados que as tecnologias de processamento podem facilitar, mas essas não são tecnologias de *Data Science* propriamente ditas, elas dão suporte ao *Data Science*, como mostra a figura 2.3.

Figura 2.3: *Data Science* no contexto dos diversos processos relacionados a dados na organização



Fonte: (PROVOST, 2016)

Tecnologias de processamento de dados são importantes para muitas tarefas de negócios orientadas a dados, que não envolvem extrair conhecimento a partir de dados ou tomar decisões orientadas a eles, como processamento eficiente de transações, processamento moderno de sistemas web e gerenciamento de campanhas publicitárias online.

Por vezes o *Data Science* é confundido com o termo Data Mining. Segundo (PROVOST, 2016), o "*Data Science* é um conjunto de princípios fundamentais que norteiam a extração de conhecimento a partir dos dados, já o Data Mining é a extração de conhecimento a partir deles, por meio de tecnologias que incorporam esses princípios".

O *Data Science* é responsável por oferecer ferramentas, processos e técnicas para acompanhamento do ciclo de vida do processo de descoberta de conhecimento e geração de valor dentro das empresas (CHANG; GRADY, 2019).

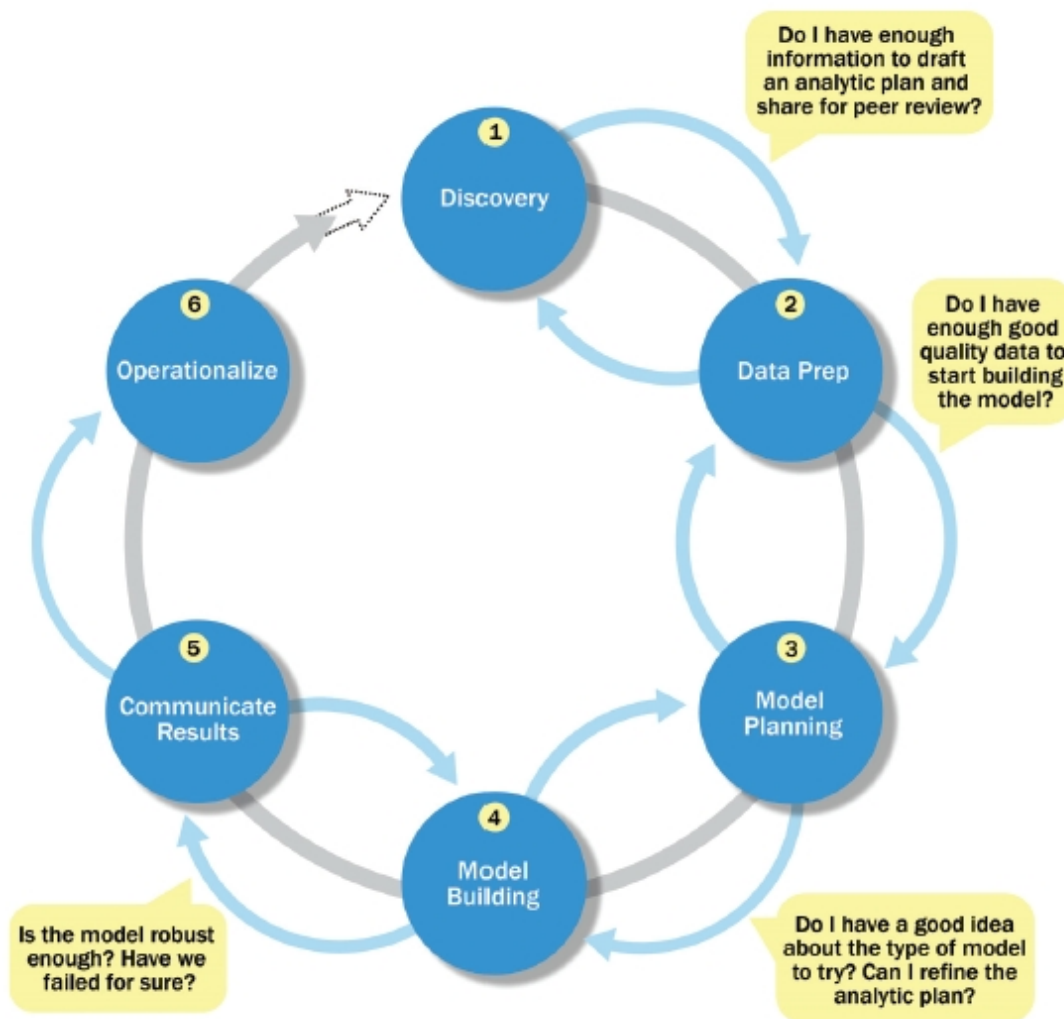
2.3 Ciclo de Vida do *Data Science*

O processo de mineração de dados envolve muita ciência e tecnologia. Segundo estudo realizado por (MOURA, 2018), onde foram avaliados diversos métodos de mineração de dados, entre eles, KDD(*knowledge discovery in databases*), CRIPS-DM, SEMMA (*Sample, Explore, Modify, Model e Assess*), DAL(*Data Analytics Lifecycle*) e SAS *Analytical Life Cycle*, pode se observar que 44% das empresas aplicam processos diversificados para análise dos dados, sendo 62,16% afirmarem que possuem algum processo de análise de dados.

2.4 *Data Analytics Lifecycle*

O *Data Analytics Lifecycle* foi originalmente desenvolvido para o curso *Data Science e Big Data Analytics da EMC* por David Dietrich, apresentado em outubro de 2012 na universidade de *Harvard business Review*. (DIETRICH D.; HELLER, 2015). O *Data Analytics Lifecycle* foi designado a resolver problemas relacionados a projetos que envolvam *Big Data* e *Data Science*. O ciclo de vida tem seis fases, que por hora ocorrem juntas, e podem movimentar-se para frente conforme o projeto avança, ou e podem se mover para trás, à medida que novas informações são descobertas e os membros da equipe aprendem mais sobre os vários estágios do projeto. O ciclo de vida tem por objetivo definir práticas recomendadas para o processo de análise desde a descoberta até a conclusão do projeto. A figura 2.4 apresenta o ciclo de vida do *Data Analytics Lifecycle*, seguido da apresentação das fases do ciclo de vida do DAL.

Fase 1 - Descoberta (*Discovery*): Esta fase consiste em aprender o domínio do negócio, incluindo o histórico da organização em forma de dados e análise. Também são avaliados os recursos disponíveis para apoiar no projeto em termos de pessoas, tecnologia, tempo e dados. É importante nesta fase estabelecer o problema do negócio, com o viés analítico, que poderá se utilizado nas fases subsequente e na formulação inicial hipóteses, para testar e começar a aprender os dados.

Figura 2.4: Ciclo de Vida *Data Analytics Lifecycle*

Fonte: (DIETRICH D.; HELLER, 2015)

Fase 2 - Preparação dos Dados (*Data Preparation*): Esta fase é utilizada para que a equipe possa extrair dados de suas fontes, aplicando métodos ETL (*extract, transform e load*), assim organizando, tratando e armazenando os dados para posteriormente realizar análises durante o projeto. Também a equipe precisa se familiarizar com os dados minuciosamente e tomar medidas para condicionar os dados.

Fase 3 - Planejamento do Modelo (*Model Planning*): Esta fase objetiva solicitar que a equipe determine os métodos, técnicas e fluxo de trabalho que pretende seguir para a fase subsequente de construção do modelo. Nesta fase também ocorre a exploração dos dados, suas relações entre variáveis, e por consequência a seleção dos modelos mais adequados para atender o modelo.

Fase 4 - Construção do Modelo (*Model Building*): Nesta fase a equipe constrói e executa modelos com base na atividade de planejamento do modelo, aplicando testes,

treinamentos dos modelos. Aplica-se também uma análise se as ferramentas e o ambiente previsto será robustos o bastante para as execuções dos modelos.

Fase 5 - Comunicar os Resultados (*Communicate Results*): Esta fase visa identificar e apresentar em colaboração a todas as partes interessadas os resultados do projeto, relacionando ao objetivo do projeto desenvolvido na fase 1. A equipe deve apresentar suas descobertas, quantificar o valor do negócio e desenvolver breve aos interessados do projeto.

Fase 6 - Operacionalizar (*Operationalize*): Esta fase abrange as entregas finais do projeto, como relatórios, finais, sínteses, códigos fontes, documentação técnica. Também é executado um projeto piloto para implementar os modelos em um ambiente de produção.

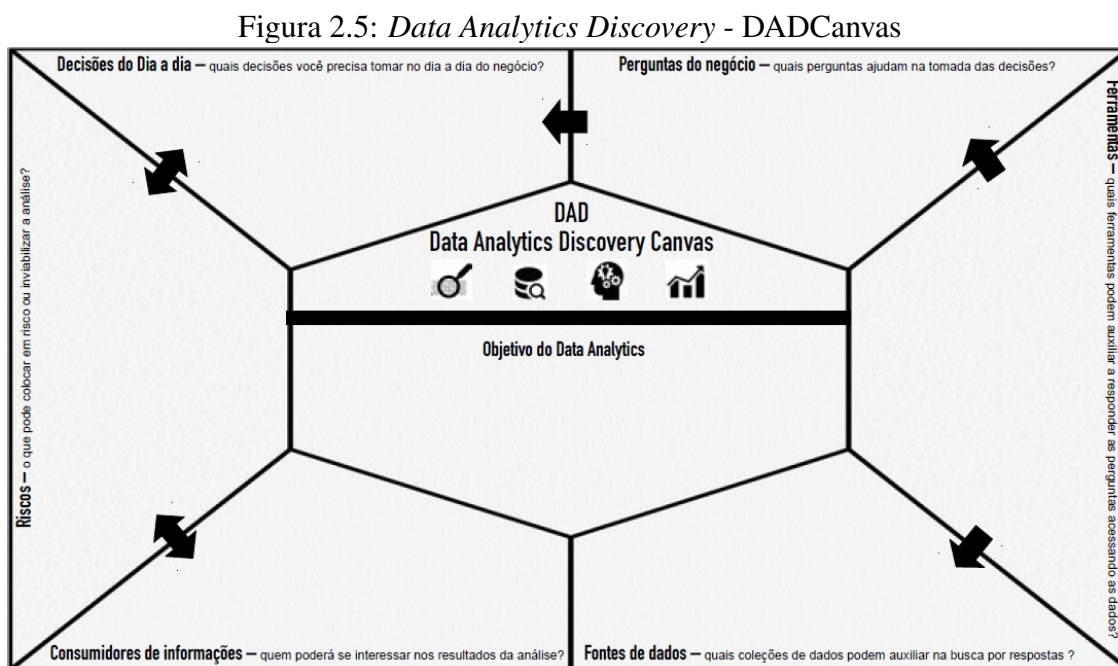
2.5 DADCanvas

Dentro do ciclo do *Data Analytical Lifecycle* existe uma etapa de descoberta (Discovery) utilizada para entendimento do problema de negócio que o projeto de dados visa atender. Esta é uma etapa crítica para sucesso do projeto. Para isto, foi criado o *Data Analytics Discovery Canvas* (DADCanvas), desenvolvido e aplicado por (MOURA, 2018). O DADCanvas teve sua primeira versão criada no ano de 2018 e em 2019 teve uma pequena reestruturação, para torna-lo visualmente mais atrativo e de uma melhor compreensão pelos usuários. O DADCanvas foi originalmente criado para atender os problemas de compreensão dos objetivos de negócios atrelados aos objetivos de análise de dados, auxiliando assim de forma visual e resumida a estruturação dos projetos de análises de dados (MOURA, 2018). Foi utilizado a fase 1 (descoberta) do método DAL, e foi dividido em sete partes, são elas:

- Fonte de dados: área destinada para registrar as fontes dos dados, ou seja, listar as fontes de dados internos e externos à empresa que possam auxiliar na composição da análise.
- Perguntas do negócio: área destinada para registrar os potenciais problemas de negócio que o projeto busca responder.
- Consumidores de informação: área destinada para registrar os *stakeholders* do projeto, as pessoas que tem interesse nos resultados obtidos com a aplicação da análise e que possivelmente irão tomar decisões baseados nelas.

- Decisões do Dia a dia: área destinada para registrar as decisões que são necessárias tomar no dia a dia, para responder perguntas de negócio.
- Objetivo do *Data Analytics*: área destinada para registrar o objetivo geral da análise, o problema central que se busca resolver.
- Ferramentas: área destinada para registrar as ferramentas necessárias que podem auxiliar a responder as perguntas acessando os dados.
- Riscos: área destinada para registrar riscos potenciais ao projeto, como problemas dados insuficientes para a análise, qualidade dos dados coletados, conhecimento da equipe, tempo e custos vinculados ao desenvolvimento do projeto.

O preenchimento do DADCanvas segue uma ordem, de acordo com o encadeamento dos pré-requisitos: 1) Decisões do Dia a dia; 2) Perguntas do negócio; 3) Fontes de Dados; 4) Consumidores de informação; 5) Ferramentas; 6) Riscos; 7) Objetivo do *Data Analytics*. A figura 2.5 apresenta o DADCanvas.



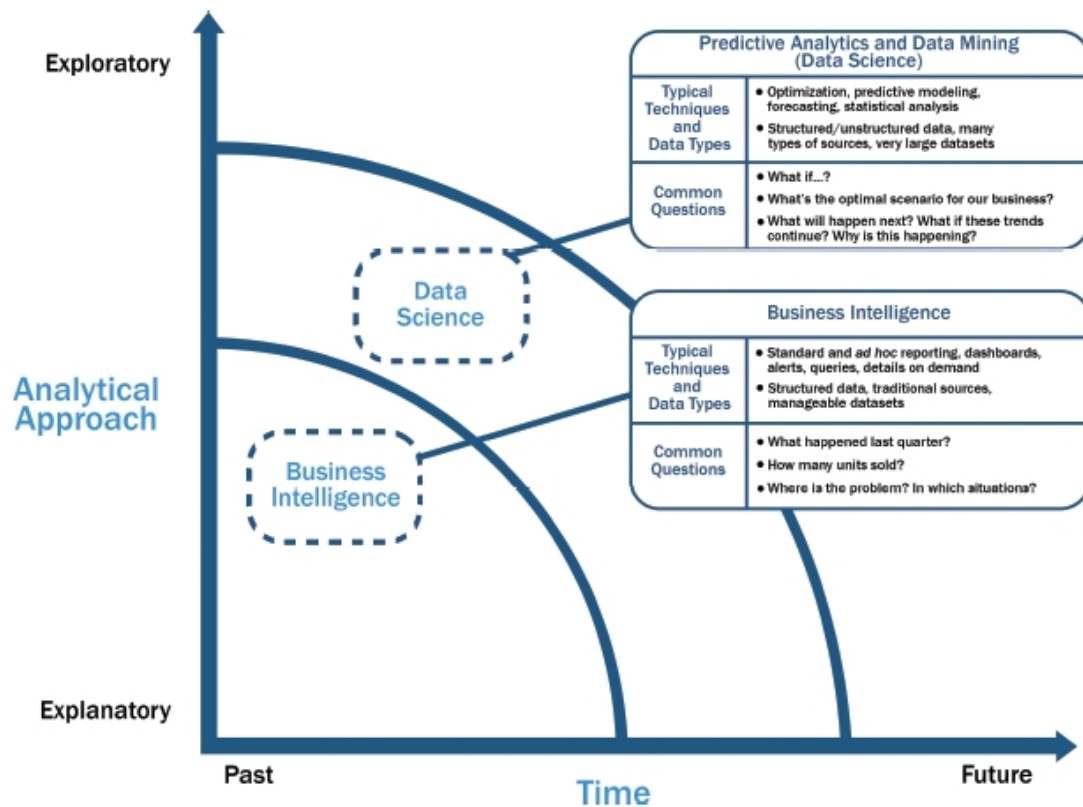
Fonte: (MOURA, 2018)

2.6 Business Intelligence

Apesar de muito conteúdo sobre análise de dados, é importante distinguir a diferença entre BI (*Business Intelligence*) e *Data Science*. A figura 2.6, compara os grupos de

técnicas analíticas de cada grupo.

Figura 2.6: Comparativo *Business Intelligence* e *Data Science*



Fonte: (DIETRICH D.; HELLER, 2015)

Diversos são as suas características, mas uma maneira de avaliar o tipo de análise é interpretando o horizonte de tempo e o tipo de abordagem analítica apresentado na figura 2.6. O BI fornece relatórios, painéis e consultas do negócio para o período atual ou passado, normalmente respondendo perguntas sobre dados históricos realizados, respondendo perguntas como "Quando e Onde" os eventos aconteceram.

Por sua vez, o *Data Science*, tende a usar dados desagregados, com foco na análise do presente e permitindo análises exploratórias sobre um possível comportamento dos dados presentes e históricos, mas com visão no futuro.

Outra evidência de uso de BI, é que normalmente a modelagem dos dados são estruturadas, em forma de linhas e colunas em bancos relacionais, o que pode mas não é comum esta estrutura em projetos de *Data Science*, pois a grande maioria dos dados são de fontes diferentes, não estruturados, grande volumes de dados.

É importante ressaltar que com o avanço tecnológico, muitas funções de análises

predictivas como, *Scripts R, Python Analises* e elementos gráficos de vizinhos próximos, foram incorporadas em software de BI, como visto na solução da *Microsoft Power BI desktop*.

Segundo (BONEL, 2019), o *Microsoft Power BI*, é hoje líder de mercado neste segmento, desde 2015. Nele é possível desenvolver análises, relatórios e *dashboards*, apresentando de forma gráfica e de fácil interpretação, as respostas para as perguntas de negócios, ajudando assim na tomada de decisão. Também se tornou muito popular pela baixa curva de aprendizado e proximidade visual com o *Microsoft Excel*, muito disseminado nas grandes e médias organizações, pois não é necessário ser profissional da área de tecnologia da informação para utilizá-lo.

3 METODOLOGIA

Para o desenvolvimento deste trabalho foi realizada uma pesquisa aplicada de caráter exploratório, que visa gerar conhecimentos a partir de uma aplicação prática para um problema específico de uma empresa metalúrgica. Ela está classificada como pesquisa exploratória, pois será realizada para validar o uso de *dashboard* na área de PCP de uma empresa metalúrgica.

“Este tipo de pesquisa tem como objetivo proporcionar maior familiaridade com o problema, com vistas a torná-lo mais explícito ou a construir hipóteses. A grande maioria dessas pesquisas envolve: (a) levantamento bibliográfico; (b) entrevistas com pessoas que tiveram experiências práticas com o problema pesquisado; e (c) análise de exemplos que estimulem a compreensão”.(GIL, 2002)

Para o desenvolvimento do trabalho foi necessário um estudo sobre o processo que guiará a construção do *dashboard*, o *Data Analytics Lifecycle* em suas etapas e a ferramenta do *Microsoft Power BI desktop*, bem como extração dos dados no banco de dados *Oracle SQL*. O trabalho foi desenvolvido seguindo as 6 fases do *Data Analytics Lifecycle*.

3.1 Fase 1 - Descoberta (*Discovery*)

Para esta etapa é necessário explorar e identificar perguntas de negócios, hipóteses, bases de dados e tecnologias e pessoas para suportar o desenvolvimento do *dashboard*. Nesta etapa também foi aplicado o DADCanvas sugerida pela autora (MOURA, 2018). Este modelo de apresentação, visa auxiliar na compreensão dos objetivos de negócios atrelados aos objetivos de análise de dados, bem como auxiliar de forma visual e resumida a estruturação do projeto de análise de dados. Para a real validação do problema, foi aplicado a metodologia do 5 Why's.

3.2 Fase 2 - Preparação dos Dados (*Data Preparation*)

Nesta etapa é necessário a criação dos *scripts* de extração dos dados, aplicando métodos ETL's (*Extract, transform e load*). Nesta etapa foi utilizado as informações obtidas no DADCanvas como, as fontes de dados que serão utilizadas e ferramentas que irão auxiliar neste processo aplicando o tratamento dos dados, seu armazenamento, e a familiarização com os dados, para posteriores análises.

3.3 Fase 3 - Planejamento do Modelo (*Model Preparation*)

Nesta fase é necessário avaliar as bases de dados extraídos e o objetivo do problema, identificando assim a técnica analítica a ser utilizada para a viabilização do objetivo do negócio. Nesta fase foi aplicado uma visualização “Tabela” do *Microsoft Power BI*, o que propiciou avaliar os dados extraídos, suas formatações, apresentações, e por fim avaliação do objetivo do projeto.

3.4 Fase 4 - Construção do Modelo (*Model Building*)

A análise de dados se dará de forma quantitativa para desenvolvimento dos gráficos e indicadores do *dashboard*, usando a ferramenta *Microsoft Power BI desktop*. Com a visualização "Tabela" executada na fase 3, foi possível aplicar diversos testes, aprimorando assim as visualizações a serem apresentadas aos *stakeholders*, como gráficos, filtros dinâmicos e até mesmo a performance do extrator.

3.5 Fase 5 - Comunicar os Resultado (*Communicate Results*)

A comunicação dos resultados foi realizada através de reuniões com todos os envolvidos, de forma *on-line*, ou presencial. A importância da comunicação dos resultados obtidos pela equipe, é de suma importância para o alcance dos resultados do projeto e objetivos do problema.

3.6 Fase 6 - Operacionalizar (*Operationalize*)

Nesta fase é necessário efetuar as entregas do projeto, como relatórios, *scripts*, e documentação técnicas. Nesta fase foi efetuado uma reunião com a diretoria, formalizando a entrega e abrindo discussões sobre a implantação do projeto.

4 APRESENTAÇÃO DO *DASHBOARD* DE GESTÃO

A necessidade de informações mais rápidas e precisas direcionadas para as tomadas de decisão dos gestores e do corpo operacional, fez com que eu realizasse este trabalho. Inicialmente foi realizada uma reunião com a diretoria da empresa, a qual me solicitou um trabalho para agilizar as informações hoje presentes em forma de relatório no ERP Infor, para uma visão mais amigável, com filtros para possíveis análises individuais. Como o ERP Infor adquirido não possui muitos recursos gráficos, falei que iria estudar alguma solução de mercado e então apresentaria para o mesmo.

No estudo realizado sobre as possíveis ferramentas de análises visuais, e também levando em consideração que a empresa já possuía uma boa relação com os softwares da *Microsoft*, escolhi o *Microsoft Power BI desktop* como uma solução para este problema.

Com a minha experiência em extração de dados, e com a experiência do diretor em análise dos dados, juntos fizemos alguns ensaios em informações que hoje já possuíamos de forma textual, mas agora apresentando-as de forma visual. Nestes pequenos ensaios e ainda por muitas vezes poucos explorados, entendemos, que era o momento de sugerir ao time de gestão da empresa o que havíamos construído, e juntos discutirmos melhorias e área de atuação para então desenvolver o primeiro painel de gestão para um problema real da empresa.

Neste momento realizamos uma reunião com o GD (Grupo Diretivo), o qual foi apresentado os ensaios montados para a equipe. Após algumas discussões e análise, foi evidenciado que o ensaio nos mostrou um grande potencial de trabalharmos com maior agilidade na tomada de decisão, bem como explorar outras informações hoje ocultas devido a dificuldade de interpretação com os atuais dados. Nesta mesma reunião, foi decidido que a área piloto para o primeiro painel de gestão seria a área de PCP, pois é nela que concentra-se as informações pertinentes à operação da empresa, como acompanhamento da manufatura, compras de material, cronogramas de projetos entre outras.

A partir desta ideia, iniciamos o trabalho com a área de PCP, marcando uma reunião com a equipe operacional e gestão da área, onde foi executada a mesma reunião apresentada para o GD. Esta reunião teve uma boa receptividade, e nela já começamos a discutir possíveis problemas do dia a dia, e como poderíamos resolver de forma ágil e visual as informações para a tomada de decisão. Após 4 horas de reunião, e diversos *insights* foi decidido que, um dos problemas que a empresa enfrentava era a dificuldade em prever o que estava impactando o atraso de fabricação dos produtos.

A partir deste *insight*, foram realizadas outras 2 reuniões, agora presente somente o coordenador da área de planejamento, 2 membros sênior operacionais do PCP, o supervisor da área de materiais, o supervisor da área de produção, e o supervisor da área de suprimentos. Nesta reunião que durou 3 horas, foram discutidos diversos problemas que afetam a entrega do produto manufaturado. Dentre estes diversos problemas levantados, 2 problemas se sobressaíram. Um era que a matéria prima não chegava na data correta, e portanto atrasava a fabricação do item, e o outro problema era que existia uma carga excessiva de trabalho em alguns centros de trabalhos fabril, acarretando assim atrasos. Como a fundamentação teórica do DADCanvas, propõe a solução de um problema de cada vez, optamos pela escolha do problema de atraso de matéria prima, uma vez que o problema de excesso de trabalho em centros de trabalho pode ser resolvido simplesmente com a introdução de horas extras e trabalhos em dias não uteis como sábado e domingo.

Como a empresa possui um perfil de resolver problemas sem muitas teorias, somente considerando experiência, já na primeira reunião apresentei o DADCanvas para que pudéssemos juntos montar o objetivo do problema, e então seguir o trabalho fundamentado pela teoria. Apesar do seu perfil não teórico, a apresentação do DADCanvas teve uma boa aceitação pela equipe de trabalho, e logo entenderam seu princípio.

Após apresentar o DADCanvas e suas premissas, iniciamos o preenchimento do documento DADCanvas. Já na primeira reunião com a equipe, o preenchimento do DADCanvas foi executado. Porém após o preenchimento, apliquei a metodologia dos 5 Why's, para realmente validar se o problema realmente estava bem definido. Esta metodologia foi criada por Sakichi Toyoda para a *Toyota Industries Corporation*, e tem como objetivo identificar a causa raiz do problema, aplicando 5 vezes a pergunta por quê. (SERRAT, 2017) A figura 4.1, apresenta a 1º versão do DADCanvas.

Na aplicação do 5 Why's foi constatado que o problema de atraso de matéria prima é sim o motivo de muitos atrasos na produção, porém foi visto que dificilmente poderemos ter controle sobre alguns atrasos, devido ao modelo fabril de nossa empresa que é produção única para o cliente, e, portanto existem diversas alterações solicitadas pelos clientes, que são executadas ao decorrer do tempo de entrega do produto final, e fazem parte do negócio da empresa. A tabela 4.1, apresenta a 1º versão do 5 Why's.

Figura 4.1: 1º Versão do DADCanvas

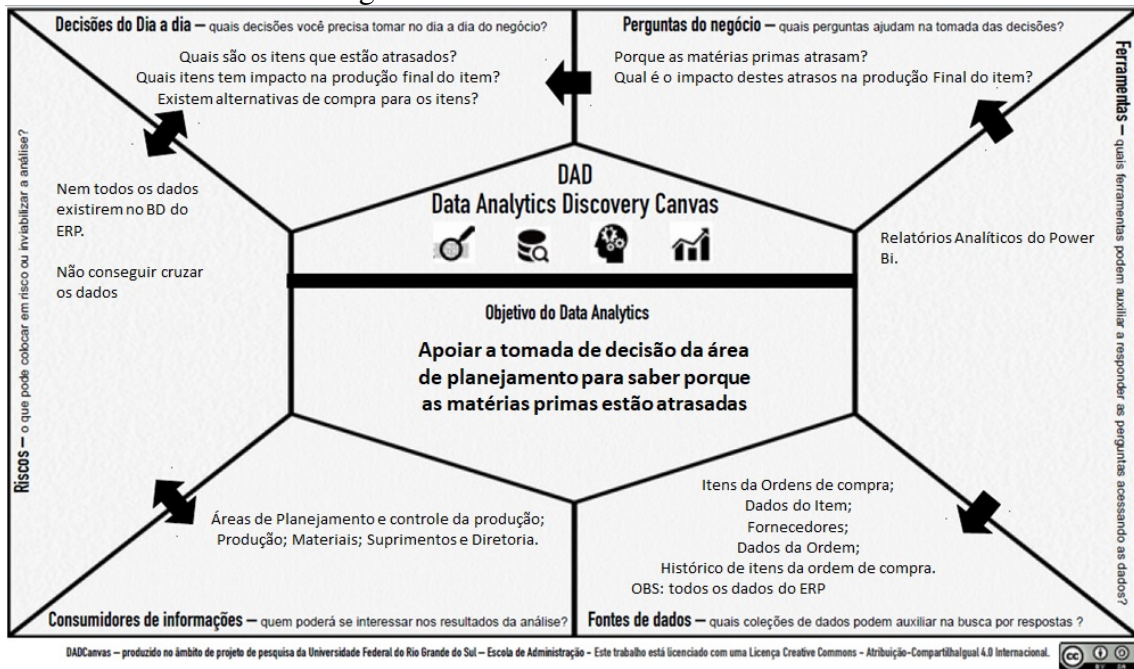


Tabela 4.1: 1º Versão 5 why's do problema

Quadro 5W

Problema: Matérias primas Atrasadas

HIPÓTESE 1

Num _{pergunta}	Pergunta	Resposta
1	Porque as matérias primas estão atrasadas?	Porque o Fornecedor Atrasou.
2	Porque o fornecedor atrasou a entrega?	Porque colocamos o pedido atrasado.
3	Porque colocamos o pedido atrasado?	Porque o prazo do projeto estava curto.
4	Porque o prazo do projeto estava curto?	Porque área comercial vendeu com prazo apertado.
5	Porque área comercial vendeu com prazo apertado?	Porque se não fosse no tempo do cliente perderia a venda.

HIPÓTESE 2

Num _{pergunta}	Pergunta	Resposta
1	Porque as matérias primas estão atrasadas?	Porque a área de planejamento não respeita o lead time na liberação das ordens.
2	Porque não respeita o lead time na liberação das ordens?	Porque são obrigados a colocar a ordem na data da necessidade.
3	Porque são obrigados a colocar a data na necessidade?	Porque precisam atender ao cronograma do projeto.
4	Porque precisão atender o cronograma do projeto?	Porque senão não entrega o produto.
5	Porque não entrega o produto?	Porque as matérias primas estão atrasadas.

HIPÓTESE 3

Num _{pergunta}	Pergunta	Resposta
1	Porque as matérias primas estão atrasadas?	Porque o tempo de processamento das ordens de compra são lentos.
2	Porque o tempo de processamento é lento?	Porque o processos de orçamento é muito pesado.
3	Porque o processo de orçamento é muito pesado?	Porque é necessário fazer 3 orçamentos.
4	Porque é necessário fazer 3 orçamentos?	Porque é necessário comprar no menor preço.
5	Porque é necessário comprar no menor preço?	Porque precisa atender ao budget de compras.

HIPÓTESE 4

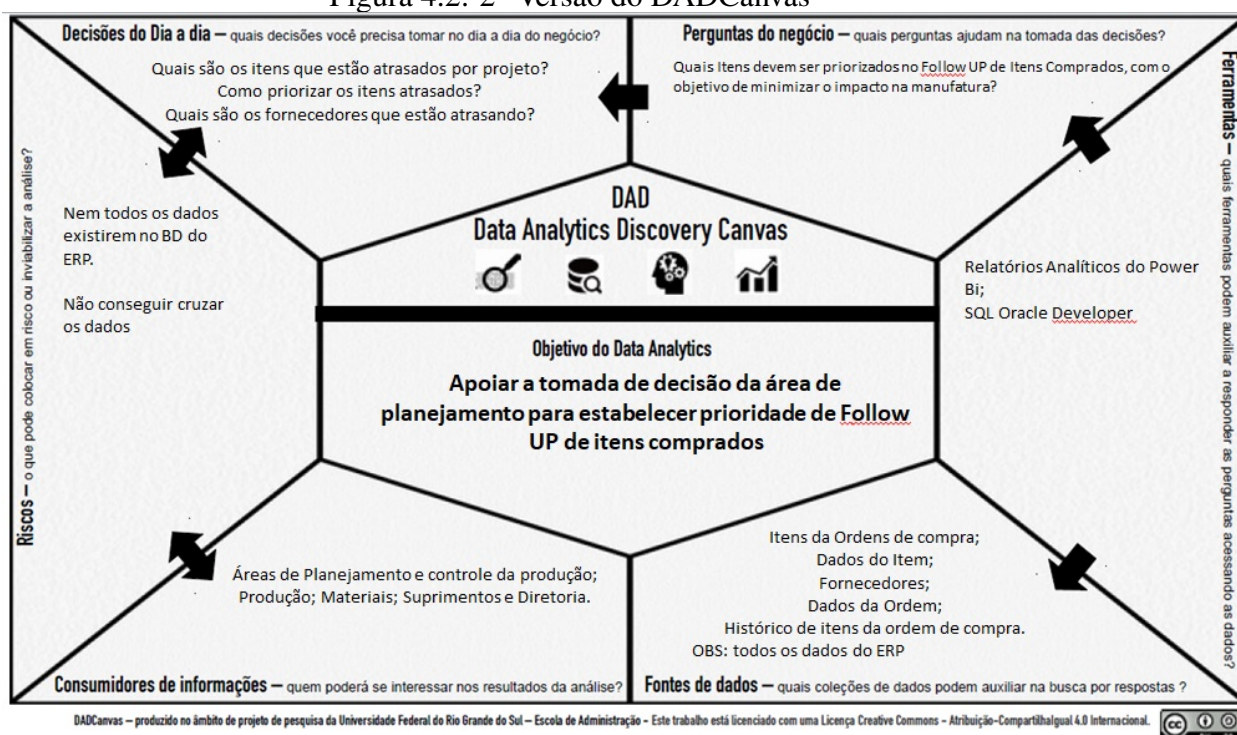
Num _{pergunta}	Pergunta	Resposta
1	Porque as matérias primas estão atrasadas?	Porque tudo precisa passar por solicitação de compras.
2	Porque tudo precisa passar por SC?	Porque não temos listas de contratos por item.
3	Porque não temos lista de preços para os itens?	Porque falta tempo para estabelecer PN para os itens.
4	Porque falta tempo para definir PN para os itens?	Porque tem muitas SC's para processar.
5	Porque se tem muitas SC's para processar?	Porque por não ter PN definido, tudo tem que ser orçado.

Fonte: Próprio Autor

Tento em vista que a primeira versão do DADCanvas não teve sucesso para aplicação, mas serviu para o aprendizado e a compreensão do problema.

Na segunda reunião realizada com a mesma equipe, os resultados e análise do 1º DADCanvas foi apresentado aos membros. Neste momento iniciamos uma nova discussão, agora entendendo que o atraso de matéria prima, fazia parte do negócio, e, portanto o foco era como tratar melhor com este critério, antes observado como problema. Após a reunião, foi refeito o DADCanvas com ênfase no impacto de cada item comprado e seu atraso. A Figura 4.2, apresenta a 2º versão do DADCanvas.

Figura 4.2: 2º Versão do DADCanvas



Fonte: Próprio Autor

O 2º DADCanvas, também foi submetido ao método 5 Why's, e o mesmo mostrou que o problema evidenciado estava sendo bem definido. A tabela 4.2, apresenta a 2º versão do 5 Why's.

Tabela 4.2: 2º Versão 5 why's do problema

Quadro 5W		
Problema:	Priorizar o Follow UP de itens comprados, para minimizando o impacto na manufatura?	
HIPÓTESE 1		
Num_pergunta	Pergunta	Resposta
1	Priorizar o Follow UP de itens comprados, para minimizando o impacto na manufatura?	Porque hoje fizemos o follow-up pela data de entrega.
2	Porque hoje fizemos o follow-up pela data de entrega?	Porque é o único dado que possuímos de prioridade.
3	Porque é o único dado que possuímos de prioridade?	Porque hoje não temos outras informações amarradas ao item para ter outra regra.
4	Porque hoje não temos outras informações amarradas ao item para ter outra regra?	Porque ao gerar a demanda de compra não amarramos o item manufaturado a compra.
5	Porque ao gerar a demanda de compra não amarramos o item manufaturado a compra?	Porque o software ERP não permite.
HIPÓTESE 2		
Num_pergunta	Pergunta	Resposta
1	Priorizar o Follow UP de itens comprados, para minimizando o impacto na manufatura?	Porque precisamos entender o que é mais importante para a empresa.
2	Porque precisamos entender o que é mais importante para a empresa?	Porque estamos fazendo follow-up por data e não por impacto na produção.
3	Porque estamos fazendo follow-up por data e não por impacto na produção?	Porque hoje não temos outras informações amarradas ao item para ter outra regra.
4	Porque hoje não temos outras informações amarradas ao item para ter outra regra?	Porque ao gerar a demanda de compra não amarramos o item manufaturado a compra.
5	Porque ao gerar a demanda de compra não amarramos o item manufaturado a compra?	Porque o software ERP não permite.
HIPÓTESE 3		
Num_pergunta	Pergunta	Resposta
1	Priorizar o Follow UP de itens comprados, para minimizando o impacto na manufatura?	Porque precisamos saber qual é o item que mais irá atrasar a produção da manufatura.
2	Porque precisamos saber qual é o item que mais irá atrasar a produção da manufatura?	Porque assim podemos estabelecer prioridade de cobrança e acompanhamento.
3	Porque precisamos estabelecer prioridade de cobrança e acompanhamento?	Porque assim conseguiremos minimizar o impacto da manufatura.
4	Porque conseguiremos minimizar o impacto da manufatura?	Pois assim estaremos dando prioridade de follow-up para os itens que são mais impactantes na manufatura.

Fonte: Próprio Autor

Com a aplicação do 5 Why's, e por ele ter validado o problema então agora foi o momento de iniciar as demais fase do DAL , pois a fase 1 de descoberta já havia sido efetuada.

A Fase 2 de Preparação dos dados foi uma das fases mais desafiadoras, e que levaram maior tempo no desenvolvimento do trabalho. Apesar da etapa de fonte de dados já estar preenchida no DADCanvas, e que por hora estamos falando que a principal fonte geradora partia de um software ERP, onde por conceito utiliza um banco de dados relacional e estruturado, foi observado que algumas informações pertinentes não existiam nas fontes de dados. Apesar da empresa trabalhar com o MRP (Material Requirement Planning), e o mesmo possuir as informações de fornecimento e demanda, a estratégia de abastecimento que é compra para estoque, e não compra para projeto, criou uma dificuldade na extração dos dados , pois não existia a relação da ordem de compra (fornecimento) x demanda (Ordem de produção – Item manufaturado) x Projeto.

Devido a esta dificuldade, porém de extrema importância para a tomada de decisão, foi necessário solicitar ao desenvolvedor do ERP Infor, um script que montasse esta relação de fornecimento x demanda x projeto, e os armazenasse no banco de dados. O script é executado via JOB (conjunto de operações que se executam simultaneamente), e utiliza como base, as informações armazenadas pelo processo do MRP, que por sua vez é executado a noite devido ao grande volume de dados que necessita ler e processar para encontrar as relações de fornecimento x demanda. Após algumas semanas de desenvolvimento do script, testes e validações das informações geradas, então foi possível criar a relação necessária para a extração dos dados para o *Dashboard*.

As demais informações pertinentes para o painel como : itens da ordens de com-

pra, dados do item, cadastro de fornecedores, dados da ordem de compra, e histórico de itens da ordem de compra, já se encontravam armazenadas no banco de dados, somente foi necessário extrair e tratar algumas informações.

Para a extração dos dados, foi utilizado a ferramenta *SQL Oracle Developer*. Foi criado um script para relacionar os dados necessários para as fases posteriores. Diversas foram as aplicações das ETLs, até chegar a um exemplar de informações pertinentes para possíveis análises de dados no *Microsoft Power BI*.

A versão final do *script* de extração dos dados apresentou um algoritmo de 427 linhas com mais 3 funções Oracle, criadas para performar a extração. Foram necessários aproximadamente 150 horas de desenvolvimento para a solução final.

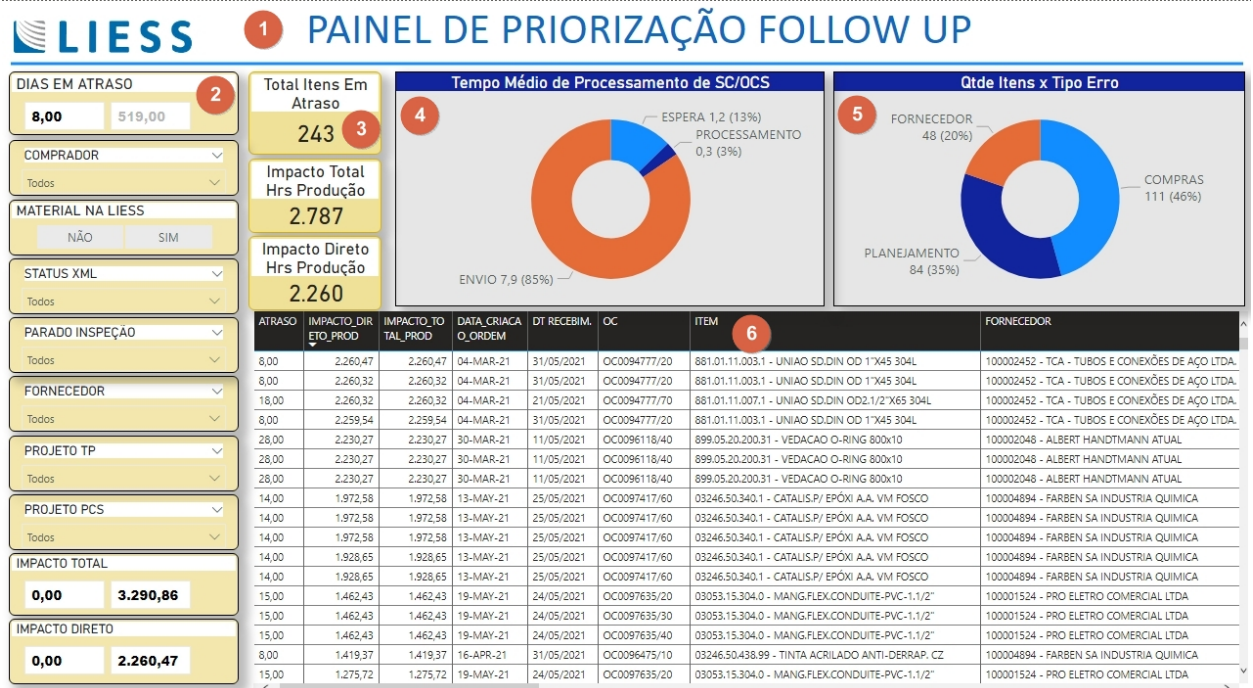
Na fase 3 do DAL (Planejamento do Modelo), foi aplicado uma análise descritiva, objetivando de forma dinâmica entender como os dados se organizam e seu comportamento. O *script* desenvolvido na fase de preparação de dados, foi aplicado no *Microsoft Power BI desktop*. Por utilizar um aplicativo local e não na nuvem, foi possível aplicar o *Direct query do Power BI*, assim não foi necessário armazenar os dados em banco de dados, efetuando a consulta dos dados online, apenas com a execução do atualizar informações. Como o trabalho realizado foi a âmbito experimental, esta metodologia é aplicável.

Após importar o *script* para o *Microsoft Power BI* e atualizar as informações, foi utilizado o tipo de visualização "Tabela", o que consiste mostrar todos os dados em forma de colunas e linhas. Inicialmente esta visualização facilita a equipe na identificação dos métodos, técnicas analíticas e fluxo de trabalho, fornecendo assim condições para explorar os dados, suas relações e então validar se o modelo atenderá ao problema negócio.

Na fase 4 de construção do modelo, foi aplicado o aprendizado dos conhecimentos dos dados para a geração do *dashboard* de gestão. Nesta fase foi desenhado um layout personalizado para a empresa, bem como um menu de filtros. Estes filtros foram desenvolvidos pensando em agilizar a centralização das informações pertinentes para a tomada de decisão da equipe.

Para um melhor entendimento do modelo proposto do *dashboad* de gestão, seguem figuras do resultado final do modelo. A figura 4.3, representa todo o painel de gestão criado para os *stakeholders*. Nele foi aplicado o script de extração de dados desenvolvido na fase 2. O painel apresenta as seguinte características, e está dividido e 6 grandes blocos de informação.

Figura 4.3: Dashboard de Gestão



Fonte: Próprio Autor

Bloco 1 – Neste bloco é apresentado o nome do painel, e o que ele representa. Figura 4.4.

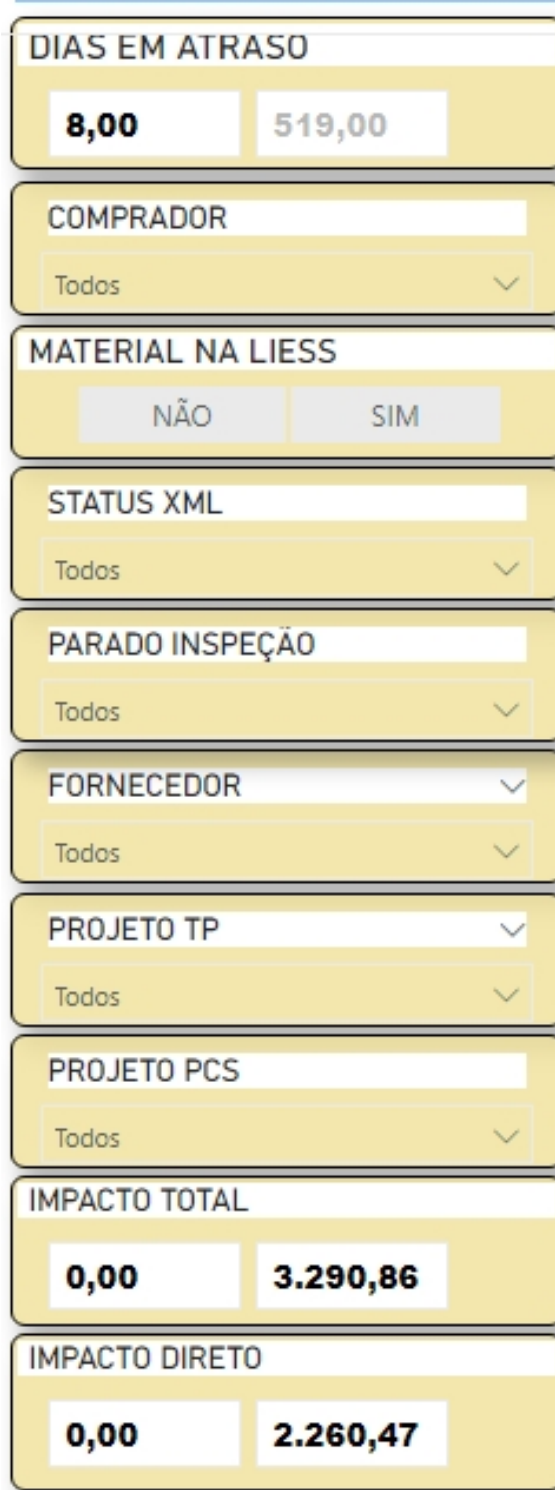
Figura 4.4: Cabeçalho do Painel



Fonte: Próprio Autor

Bloco 2 – Neste bloco, representado pela figura 4.5 é apresentado todos os possíveis filtros aplicáveis aos dados apresentados. É importante salientar que qualquer filtro aplicado, afetará nas informações apresentadas. Os Filtros desenvolvidos são:

Figura 4.5: Filtros do Painel



The image displays a vertical stack of ten filter panels. Each panel has a title and a selection area. The first and last panels show numerical values in two boxes. The second, fourth, fifth, sixth, seventh, and eighth panels are dropdown menus. The third panel has two radio buttons. The ninth and tenth panels show numerical values in two boxes.

Filter Name	Value 1	Value 2
DIAS EM ATRASO	8,00	519,00
COMPRADOR	Todos	
MATERIAL NA LIESS	NÃO	SIM
STATUS XML	Todos	
PARADO INSPEÇÃO	Todos	
FORNECEDOR	Todos	
PROJETO TP	Todos	
PROJETO PCS	Todos	
IMPACTO TOTAL	0,00	3.290,86
IMPACTO DIRETO	0,00	2.260,47

Fonte: Próprio Autor

- DIAS EM ATRASO: Este filtro apresenta a quantidade de dias em atraso que um determinado item possui com relação à data atual. Esta data atual é referenciada pela atualização dos dados na consulta do painel.
- COMPRADOR: Este filtro apresenta o nome do comprador responsável pela gestão

do atraso do item. Esta informação é parte integral da consulta dos dados do Item ordem de compra.

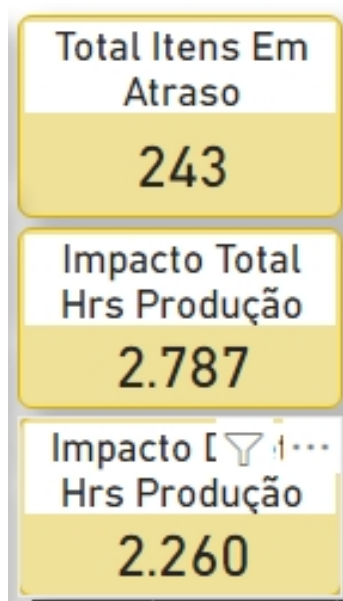
- **MATERIAL NA EMPRESA:** Este filtro visa eliminar da consulta, os itens que se encontram em posse da empresa, porém ainda não foram executados os procedimentos internos no sistema de contabilização. Este filtro é importante, pois pode eliminar possíveis análises errôneas de cobranças de atrasos. Todavia, também é importante saber os itens que já estão na empresa para então cobrar agilidade de processamento interno da entrada de pedidos.
- **STATUS XML:** Este filtro visa apresentar os itens os quais os fornecedores emitiram Nota Fiscal Eletrônica (XML), portanto o item está em via de transporte. Esta informação pode auxiliar na identificação dos itens que mesmo estando atrasados, e impactantes, já que estão em rota de transporte.
- **PARADO NA INSPEÇÃO:** Um item que está parado na inspeção, significa que o mesmo está sobre posse da área da qualidade, aguardando processo de análise da área. Por sua vez, o item que está neste status, não libera o abastecimento do item para a manufatura e portanto impacta no problema em questão.
- **FORNECEDOR:** Este filtro apresenta os fornecedores que estão com itens em atrasos. Ele tem relevância quando o painel é utilizado pela área de suprimentos, assim podem efetuar cobranças dos atrasos por comprador e fornecedor.
- **PROJETO TP:** Este filtro tem uma nomenclatura própria. Como a empresa trabalha em manufatura para projeto, faz-se necessário o uso de projeto Pai. Este projeto pai é chamado de Projeto TP. A finalidade do projeto TP, é agrupar diversos projetos menores. Toda venda efetuada na empresa, transforma-se em um número de projeto TP, e é a partir dele que temos todas as informações de entrega ao cliente, e por consequência a entrega dos itens manufaturados.
- **PROJETO PCS:** Este filtro apresenta projetos menores que o projeto TP. Um projeto TP possui 1 ou mais projetos PCS's, que ao final da entrega de todos os projetos PCS's, o projeto TP é concluído e entregue para o cliente.
- **IMPACTO TOTAL:** Este filtro representa o impacto em horas total de produção, que um determinado item atrasado, poderá impactar na produção total da entrega do item manufaturado. Este número representa todos os níveis de produção, ele indicará o quando deixo de produzir em toda cadeia de produção, por falta desta matéria prima.

- **IMPACTO DIRETO:** Este filtro representa o impacto em horas total de produção, que um determinado item atrasado, poderá impactar na produção total da entrega do item manufaturado direto. Este numero representa somente o nível superior da estrutura do produto, e indicará o quando deixo de produzir diretamente, por falta desta matéria prima.

Bloco 3 – Indicadores de grandezas – Neste bloco, representado pela figura 4.6 são apresentados 3 indicadores de grandezas, aplicados aos filtros. Os indicadores de grandezas são:

- **Total de itens atrasados:** Apresenta a quantidade total de itens atrasados. Nele é aplicado a regra de que atraso é sempre (data atual – data prevista recebimento). Caso este calculo retorno um valor negativo, então este fará parte da contagem de itens em atraso.
- **Impacto Total Hrs Produção:** Apresenta a quantidade de horas de produção total afetada pelo atraso de um componente. Este indicador é muito relevante, quando os filtros de projetos são aplicados, pois apresentam o risco do abastecimento na entrega do projeto.
- **Impacto Direto Hrs Produção:** Apresenta a quantidade de horas de produção direta afetada pelo atraso de um componente. Este indicador é muito relevante, quando os filtros de projeto PCS é aplicados, pois apresentam o risco do abastecimento na entrega do projeto PCS para com o Projeto TP.

Figura 4.6: Indicadores de Grandezas



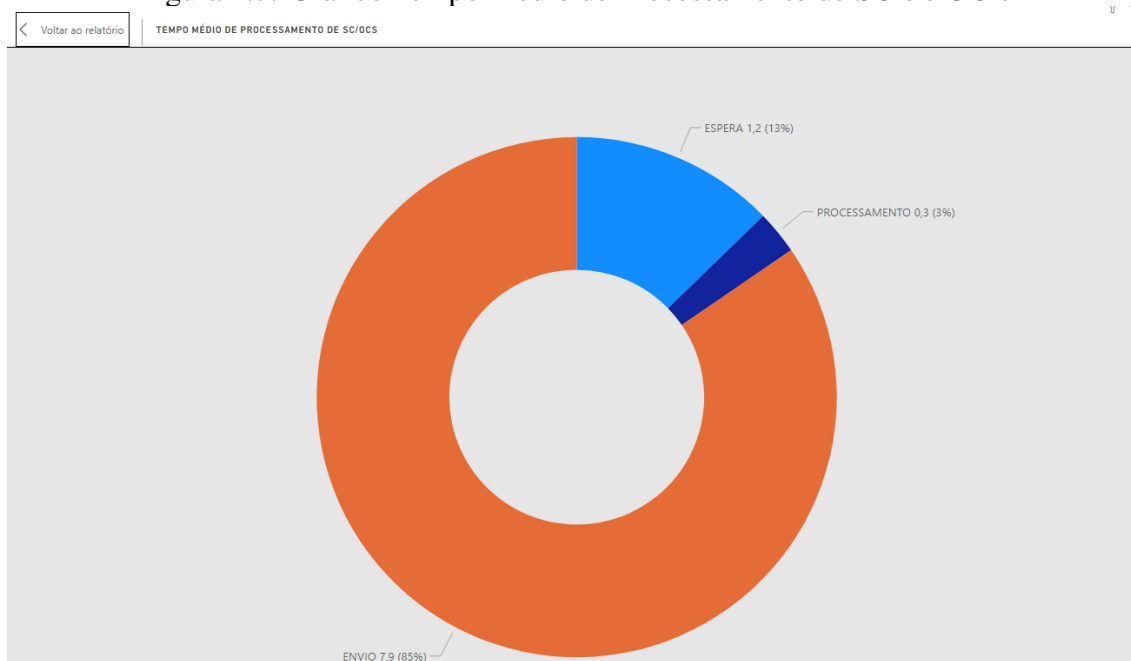
Fonte: Próprio Autor

Bloco 4 – Gráfico Tempo Médio de Processamento de SC/OC – Este gráfico, represen-

tado na figura 4.7, foi criado para medir o tempo médio de processamento interno das SC's (Solicitação Cotação) e OC's (Ordens de Compra). Esta medição pode auxiliar a área de suprimentos a averiguar possíveis distorção no tempo de processamento, pois por experiência, o tempo médio de processamento é de 5 dias. A medição do tempo total de processamento é realizado pelos tempos:

- Espera (Período que nenhum usuário mexeu no documento no sistema)
- Processamento (Período em que foi enviado o pedido, e aguarda alguma autorização de aprovação interna da compra)
- Envio (Período em que foi selecionado o fornecedor, enviado o pedido, e aguarda retorno de confirmação).

Figura 4.7: Gráfico Tempo Médio de Processamento de SC's e OC's



Fonte: Próprio Autor

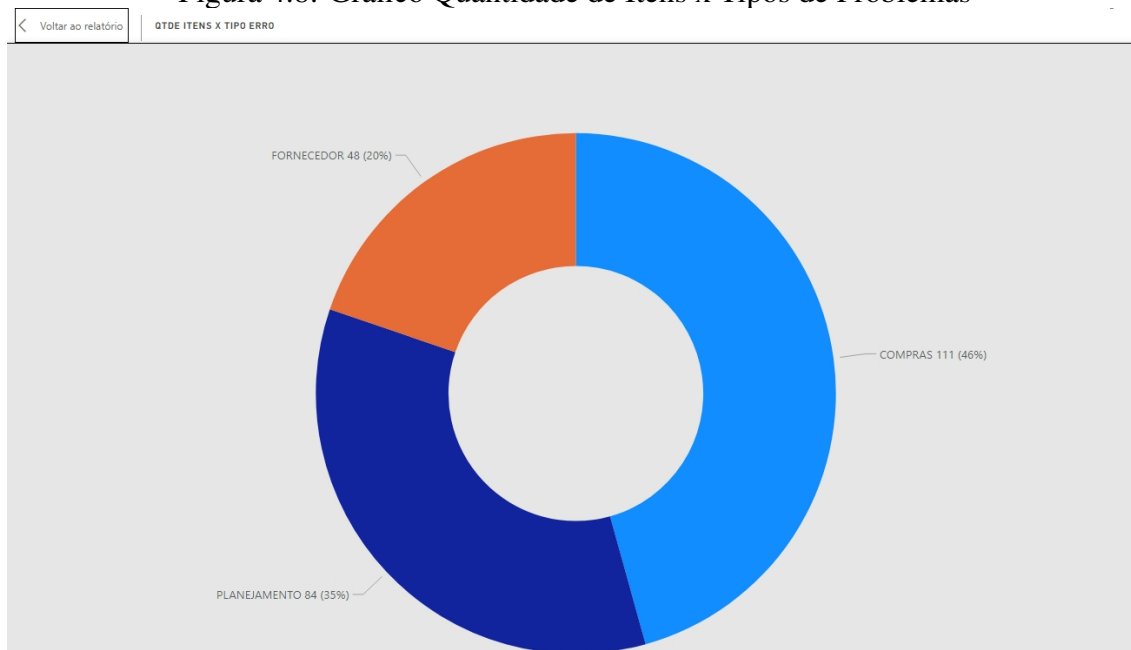
Bloco 5 – Qtde Itens x Tipo de Problema – Este gráfico, representado pela figura 4.8, foi criado para identificar potenciais erros que acarretam no atraso de um determinado item. Sua medição dar-se:

- **ERRO PLANEJAMENTO:** é sinalizado quando um determinado fornecimento é gerado em desconformidade com o tempo de fornecimento do item + o tempo de transporte + tempo médio processamento pedido: Exemplo: Um fornecimento para 10/06/2021. Porém o tempo de fornecimento = 30, tempo transporte = 5, tempo processamento pedido = 5. Cálculo: 10/06/2021 – 40 = 01/05/2021 – Caso o se-

tor de planejamento libere a ordem de compra depois do dia 06/05/2021, então é classificado como ERRO PLANEJAMENTO.

- **ERRO COMPRAS:** Após o planejamento liberar o pedido de compra, então o setor de suprimentos tem 5 dias para processar o pedido. Caso ele não libere neste período, então é classificado como ERRO COMPRAS.
- **ERRO FORNECEDOR:** Todos os demais erros são classificados como ERRO FORNECEDOR, pois estavam dentro do prazo e não foi atendido pelo fornecedor.

Figura 4.8: Gráfico Quantidade de Itens x Tipos de Problemas



Fonte: Próprio Autor

Bloco 6 – Detalhes dos Pedidos de Compra – Neste bloco, representado pela figura 4.9, são apresentados mais de 25 informações pertinentes ao pedido de compra que auxiliam o usuário no direcionamento do seu trabalho para priorizar o *follow-up* de compras, podendo extrair os dados para o *Microsoft Excel* e assim enviar para áreas de interesses, bem com o fornecedor.

Na construção do *dashboard*, 3 foram as informações pertinentes para o atendimento do objetivo do DADCanvas. A informação de Atraso, Impacto Direto Horas, Impacto Total de Horas. Segue explicação detalhado do conceito de cada informação pertinente:

- **Atraso** – esta informação é medida em numero de dias em atraso. Caso o cálculo retorne valor negativo, o item esta em atraso. Esta medida é feita pelo cálculo (data

Atual – data prevista de recebimento).

- **Impacto Direto Horas** – Esta informação é media em numero de horas. Este cálculo só é executado se o atraso for > 0 . Para chegar a este medida, é aplicado uma leitura de estrutura de engenharia e então lido a quantidade de horas de produção o qual o nível superior ao item atrasado encontra-se. A tabela 4.3 apresenta um exemplo. Dado o exemplo, onde o item atrasado é o código “666”, no nível 3, o sistema percorre para cima da estrutura de engenharia do item manufaturado relacionado ao fornecimento. O primeiro item manufaturado menor que o nível da busca é o impacto de horas direta. Neste exemplo, ao executar a função o sistema iria encontrar o item “777” nível dois e apresentaria o resultado de 3 horas.
- **Impacto Total Horas** – Esta informação é media em numero de horas. Este cálculo só é executado se o atraso for > 0 . Para chegar a este medida, é aplicado uma leitura de estrutura de engenharia e então lido a quantidade de horas de produção o qual o nível 0 da estrutura é alcançado. A tabela 4.3 apresenta um exemplo. Dado o exemplo, onde o item atrasado é o código “666”, no nível 3, o sistema percorre toda a estrutura de engenharia do item manufaturado relacionado ao fornecimento, somando todas as quantidade de horas de produção. O resultado desta soma é o impacto total de horas. Neste exemplo, ao executar a função o sistema iria encontrar o item “777”, ”888” e “999”, representados pelas horas “3”+”5”+”15”. Neste exemplo o resultado apresentado seria de 23 horas.

Tabela 4.3: Exemplo de uma Estrutura de Engenharia de um produto manufaturado

NIVEL	ITEM	TIPO DE ITEM	DESCRIÇÃO	HORAS PRODUÇÃO
0	999	MANUFATURADO	BALANÇA	15
1	888	MANUFATURADO	CHAPA PRENSADA	5
2	777	MANUFATURADO	CHAPA CORTADA	3
3	666	COMPRADO	CHAPA	0

Figura 4.9: Tabela Detalhes

Voltar ao relatório														
ATRASSO	IMPACTO DIR ETO_PROD	IMPACTO TO TAL_PROD	DATA CRACA O_ORDEM	DT RECEBIM.	OC	ITEM	FORNECEDOR	COMPRADOR	UNID	FA				
42.00	0.00	0.00	27-JAN-21	27/04/2021	OC0093600/10	2670020 899.05.00.667.85 - MANIFOLD CLARIFICAÇÃO	100000288 - ALFA LAVAL LTDA	7125 - CARLOS ALBERTO DOS SANTOS	pc					
57.00	0.00	0.00	12-FEB-21	12/04/2021	OC0094135/50	2670014 899.05.00.667.92 - VALV.DUPLA SEDE 2V 2"	100000288 - ALFA LAVAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
40.00	42.46	113.04	12-FEB-21	29/04/2021	OC0094855/70	03063.58.686.2 - VALV.BORB.MAN4" 5X5 OD 304L	100000288 - ALFA LAVAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
40.00	4.54	4.54	12-FEB-21	29/04/2021	OC0094855/70	03063.58.686.2 - VALV.BORB.MAN4" 5X5 OD 304L	100000288 - ALFA LAVAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
13.00	0.00	0.00	12-APR-21	26/05/2021	OC0097638/10	03055.08.243.0 - BUJAO SX 3/8" BSP	100000394 - VALSUL VALVULAS DO SUL LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
8.00	0.00	0.00	12-APR-21	31/05/2021	OC0097716/10	03055.30.032.0 - TAMPAO 3/8" BSP CL10	100000394 - VALSUL VALVULAS DO SUL LTDA	8903 - GISELI PEREIRA VARGAS DE SOUZA	pc					
8.00	0.00	0.00	12-APR-21	31/05/2021	OC0097716/20	03055.23.003.0 - CURVA M/F 90 1/2" BSP	100000394 - VALSUL VALVULAS DO SUL LTDA	8903 - GISELI PEREIRA VARGAS DE SOUZA	pc					
36.00	0.00	0.00	15-APR-21	03/05/2021	OC0096442/10	2670012 899.05.00.202.34 - EXAUSTOR RA355	100000399 - SOLER PALAU BRASIL VENTILAÇÃO LTDA	7492 - JOAO ELVIO GARCIA	pc					
15.00	0.00	0.00	17-MAY-21	24/05/2021	OC0097480/40	CT531.22.076.3 - PROTETOR HPR-XD 220763 260A	100000451 - OKIPIRA AUTO IND E COM MAQ IND LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
15.00	0.00	0.00	17-MAY-21	24/05/2021	OC0097480/50	CT531.22.076.4 - PROTETOR HPR-XD 220764 260A	100000451 - OKIPIRA AUTO IND E COM MAQ IND LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
11.00	0.00	7.01	25-MAR-21	28/05/2021	OC0095700/10	881.16.03.048.4 - PESTAÑA OD 50.8x1.5 INOX 316L	100000526 - INTERNOX CONEXÕES E METAIS LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
15.00	0.00	0.00	20-MAR-21	24/05/2021	OC0095438/10	194.06.20.205.1 - PONTA CAME DESC.PU.V58 LAR	100000538 - IPUBRAS INDUSTRIA DE POLIURETANOS	8903 - GISELI PEREIRA VARGAS DE SOUZA	pc					
15.00	5.33	5.33	05-APR-21	24/05/2021	OC0096798/10	194.06.20.205.1 - PONTA CAME DESC.PU.V58 LAR	100000538 - IPUBRAS INDUSTRIA DE POLIURETANOS	8903 - GISELI PEREIRA VARGAS DE SOUZA	pc					
28.00	0.00	0.00	26-APR-21	11/05/2021	OC0096785/10	882.10.35.068.09 - CONE SUPERIOR 70"	100000858 - F J DA SILVA GUTERRES	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
8.00	0.00	0.00	12-APR-21	31/05/2021	OC0097640/10	194.04.20.044.1 - EIXO CX CT20-2	100000858 - F J DA SILVA GUTERRES	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
85.00	0.00	0.00	21-JAN-21	15/03/2021	OC0093798/30	CT600.10.305.2 - BICO DE CONTATO DC DIA.4,0mm	100001167 - ELETRO SOLDA COMERCIAL LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
94.00	0.00	0.00	17-FEB-21	06/03/2021	OC0094530/10	CT600.10.305.0 - BICO DE CONTATO DC DIA.3,2mm	100001167 - ELETRO SOLDA COMERCIAL LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
94.00	0.00	0.00	17-FEB-21	06/03/2021	OC0094530/20	CT600.10.305.1 - BICO DE CONTATO DC DIA.3,5mm	100001167 - ELETRO SOLDA COMERCIAL LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
94.00	0.00	0.00	17-FEB-21	06/03/2021	OC0094530/30	CT600.10.305.2 - BICO DE CONTATO DC DIA.4,0mm	100001167 - ELETRO SOLDA COMERCIAL LTDA	9001 - DIEGO TRESSOLDI DE OLIVEIRA	pc					
13.00	0.00	0.00	26-MAR-21	26/03/2021	OC0095601/10	10100090 - MATERIAIS DIV.CANTEIRO DE OBRA	100001323 - CAPPELLETTI COM E LOC DE FERRAMENTA	7492 - JOAO ELVIO GARCIA	un					
28.00	0.00	0.00	11-MAY-21	11/05/2021	OC0097345/10	LOCACAO,MAQ,EQUIP - LOCACAO MAQUINAS EQUIPAMENTOS	100001323 - CAPPELLETTI COM E LOC DE FERRAMENTA	7492 - JOAO ELVIO GARCIA	un					
31.00	0.00	0.00	07-JAN-21	08/05/2021	OC0092970/50	CONTRIB.ASSOCCL - CONTRIBUICAO ASSOCOACAO CLASSE	100001389 - ASSOC BRAS DA IND DE MAQUINAS E	9001 - DIEGO TRESSOLDI DE OLIVEIRA	un					
19.00	0.00	0.00	24-MAR-21	20/05/2021	OC0095841/10	IMOBILIZADO - IMOBILIZADO	100001507 - TORNISINOS IND E COM DE PEÇAS	9001 - DIEGO TRESSOLDI DE OLIVEIRA	un					
32.00	0.00	0.00	30-APR-21	07/05/2021	OC0096990/10	FERRAM.MODELOS - FERRAMENTAS E MODELOS	100001507 - TORNISINOS IND E COM DE PEÇAS	9001 - DIEGO TRESSOLDI DE OLIVEIRA	un					
12.00	0.00	0.00	20-MAY-21	27/05/2021	OC0097658/10	FERRAM.MODELOS - FERRAMENTAS E MODELOS	100001507 - TORNISINOS IND E COM DE PEÇAS	9001 - DIEGO TRESSOLDI DE OLIVEIRA	un					
11.00	2.33	387.72	04-MAY-21	28/05/2021	OC0097118/10	03013.20.308.0 - ESFERA CIP 369" 2" 30m3mx5m	100001514 - NOCADO DO BRASIL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
11.00	2.33	387.72	04-MAY-21	28/05/2021	OC0097118/10	03013.20.308.0 - ESFERA CIP 369" 2" 30m3mx5m	100001514 - NOCADO DO BRASIL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	pc					
15.00	1.462.43	1.462.43	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	1.271.15	1.271.15	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	0.46	1.271.61	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	1.273.98	1.273.98	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	0.46	1.274.44	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	1.274.02	1.274.02	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	0.46	1.274.47	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	1.274.10	1.274.10	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					
15.00	0.46	1.274.56	19-MAY-21	24/05/2021	OC0097635/20	03053.15.304.0 - MANG.FLEX.CONDUITE-PVC-1.1/2"	100001524 - PRO ELETRO COMERCIAL LTDA	8773 - LAURA JACQUES RIBEIRO DOS SANTOS	m					

Fonte: Próprio Autor

Para a Fase 5 do DAL, Comunicar os resultados, foi efetuado uma reunião com o diretor da empresa, a gerencia do PCP e a equipe de trabalho apresentando o resultado final do *dashboard*. Devido à pandemia do COVID-19, esta reunião não aconteceu de forma, presencial, mas sim de forma virtual. Indiferentemente do modelo de apresentação, na reunião foram apresentados os resultados obtidos com a extração dos dados, bem como a aplicação de diversos filtros para verificar a aplicabilidade dos resultados perante as problemas encontrados no negócio. A partir da apresentação do *dashboard* para os *stakeholders* alguns pontos de questionamento foram levantados, e algumas sugestões foram explicitadas. Dente os questionamentos, 4 perguntas tiveram maior ênfase, por não ter sido respondido no ato.

- Pergunta 1 : Contando com a nossa experiência, o Item “X”, não tem este impacto na produção?
- Pergunta 2 : Contando com a nossa experiência, o indicador de Qtde Itens x Tipo de erro, esta apresentando estatística errada, principalmente quando avaliamos a performance da área de planejamento. Porque ele está tão alto?
- Pergunta 3: Contando com a nossa experiência, a média de processamento de uma necessidade de compra é de 3 dias e não 8 como normalmente esta apresentando

em média.

- Pergunta 4: Contando com a nossa experiência, alguns itens apresentados, mostraram números de impacto de horas muito alto, Por quê? Apesar do uso da métrica de impacto direto produção de horas e impacto total de produção possuir uma logica correta.

Como o foco do Painel era aplicar métricas para a priorização do *follow-up* de compras, a pergunta ganhou muita discursão, inclusive criando por horas conflitos de interesses entre as áreas, onde algumas se beneficiaram com a informação e as outras tentavam se justificar. As perguntas feitas na reunião, foram registradas, e informado que seria necessário uma pesquisa com mais detalhes para explicitar as respostas, e por fim adapta-las no painel caso fosse aplicável, porem foi unanime o ponto em que este tipo de painel de gestão ajuda e muito para a tomada de decisão, pela sua agilidade em aplicação de filtros, e informações em tempo real.

A fase 6 de operacionalização foi discutida em reunião direta com a diretoria, levantando alguns pontos importantes vistos na aplicação do painel piloto. Fatores estes discutidos como: Investimento em licenças do *Microsoft Power BI 365*, compra de espaços de armazenamento, e também contratação de um profissional dedicado para a extração dos dados solicitados pelo GD.

Sobre o investimento em licenças do *Microsoft Power BI 365*, foi apresentado ao diretor da empresa um investimento em aproximadamente 10 licenças, o que acarretaria em um investimento previsto de aproximadamente R\$ 600,00 reais por mês. Foi discutido que inicialmente a empresa gostou do resultado do projeto, e que o investimento é passível de aprovação, porem foi explicitado a vontade de aplicar mais alguns *dashboards* de Gestão, agora focados em outros problema e áreas, para então iniciar a cultura de gestão visual e tomada de decisão em dados, e não mais somente contando com a experiência.

Sobre o investimento em espaço de armazenamento dos dados, foi apresentado dois cenários passíveis de um futuro mais aprimorado. Este estudo se da, pois precisamos saber a quantidade e volume de dados que seriam necessários para futuras análises e desenhos dos painéis de gestão. Como no projeto piloto foi apresentado os dados não armazenados, e sim com busca em real time pelo *Direct query do Microsoft Power BI desktop*, e então apresentados aos usuários, este modelo não necessita de armazenamento direto. Foi também explicado que este modelo tem algumas limitações, como criação de algumas métricas direta no armazenamento, mas não apresentou neste momento risco ao modelo apresentado. Ficou então definido que o investimento em armazenamento de da-

dos será uma avaliação futura, amarrada a expansão e disseminação da gestão por dados.

Sobre o investimento na contratação de um profissional dedicado para a extração dos dados, foi discutido com o diretor que propôs uma avaliação no aumento da demanda por informações nas áreas, devido à disseminação da informação, um profissional dedicado é necessário, pois caso contrário o *lead time* de entrega das demandas poderiam passar de anos, pois hoje os profissionais qualificados para este trabalho já possuem outras atividades, e a nova atividade aos profissionais acarretaria em uma demanda muito maior que o atual fornecimento. Neste ponto foi caracterizado o interesse da alta direção neste novo cargo, e portanto após o terceiro *dashboard* piloto, a empresa irá avaliar os ganhos, e se favorável, irá abrir a vaga para contratação de um profissional dedicado.

5 CONSIDERAÇÕES FINAIS

Este trabalho teve por objetivo desenvolver um projeto de *dashboard* de gestão para a área de planejamento e controle da produção, utilizando as fases do *Data Analytics Lifecycle* sendo que na Fase 1 do DAL, foi aplicado o DADCanvas para um melhor entendimento do problema a ser tratado.

Como a aplicação prática do projeto foi desenvolvido em uma empresa metalúrgica, cujo conhecimento técnico não está direcionado na tecnologia da informação, mas sim na engenharia da produção, foram necessárias buscas por bibliografia que remetessem a compreensão de termos específicos como, *Big Data*, *Data Science*, *Business Intelligence* e Ciclo de Vida dos dados, para assim sustentar a importância dos dados para a geração de informação na tomada de decisão.

Dois foram as principais dificuldades no desenvolvimento do projeto. Uma das dificuldades encontradas foi o estabelecimento do problema real, pois em reuniões para estabelecer este ponto, ficou evidente por diversas vezes, o desvio no foco do problema, não indo para o problema raiz, e sim achando que o primeiro problema era o que tínhamos que resolver. A aplicação do DADCanvas, juntamente com a metodologia 5 Why's fortaleceu e muito a solução desta dificuldade. A segunda dificuldade na execução do projeto, foi a extração dos dados, pois apesar das inúmeras informações que um software ERP gera, muitas delas não existem, sejam elas porque o software não possui, ou por decisões estratégicas que visam simplificar a operacionalização dos processos. Também ficou claro que mesmo tratando de uma extração de dados em banco de dados relacionais, muitas informações e tratamentos foram necessários para o resultado esperado para a análise.

Apesar do resultado final do trabalho ter surpreendido o Grupo Diretivo da empresa e *stackholders*, fica evidente que existe uma necessidade de aplicar algumas palestras e cursos para desenvolver a gestão e também as equipes, na solução da causa raiz do problema, uma vez que a cultura da empresa é ter um problema e resolve-lo, mas poucas vezes analisa a fundo o problema em questão.

Como trabalhos futuros, seria interessante:

- 1) Responder as perguntas executadas na apresentação do painel de gestão, para uma melhor compreensão das informações extraídas destes cenários.

- 2) Aplicar outros *dashboards* em outras áreas para disseminar a gestão por dados.

Por fim, o autor compreende que o trabalho teve seus objetivos atingidos, pois o

dashboard de gestão auxiliou a área de PCP a priorizar os itens para o *Follow-up*, bem como apresentou outras possíveis problemas antes escondidos, e que agora podem ser trabalhados em outros painéis.

REFERÊNCIAS

- ALMEIDA 2002. 2002. <<http://revista.ibict.br/ciinf/article/view/955>>. Accessed: 2021-05-25.
- ANGELONI, T. Elementos intervenientes na tomada de decisão. **Ciência da Informação**, v. 32, 04 2003.
- BONEL, c. **Power BI Black Belt: um treinamento faca na caveira através dos principais pilares de um projeto prático de business intelligence, usando o Microsoft Power Bi**. [S.l.]: Perse,São Paulo, 2019.
- CHANG, W.; GRADY, N. **NIST Big Data Interoperability Framework: Volume 1, Definitions**. [S.l.]: Special Publication (NIST SP), National Institute of Standards and Technology, Gaithersburg, MD, 2019.
- DHAR, V. Data science and prediction. **Communications of the ACM**, ACM New York, NY, USA, v. 56, n. 12, p. 64–73, 2013.
- DIETRICH D.; HELLER, B. Y. B. **Data Science Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data**. [S.l.]: John Wiley Sons, Inc., 2015.
- ERDMANN, R. H. **Organização de sistemas de produção**. [S.l.]: Insular, 1998.
- GIL, A. **Como elaborar projetos de pesquisa**. Atlas, 2002. ISBN 9788522431694. Available from Internet: <<https://books.google.com.br/books?id=X4uvAAAACAAJ>>.
- MOURA, K. V. d. Data science: um estudo dos métodos no mercado e na academia. 2018.
- PROVOST, F. F. **Data Science para Negócios**. [S.l.]: Alta Books, 2016.
- SERRAT, O. The five whys technique. In: _____. **Knowledge Solutions: Tools, Methods, and Approaches to Drive Organizational Performance**. Singapore: Springer Singapore, 2017. p. 307–310. ISBN 978-981-10-0983-9. Available from Internet: <https://doi.org/10.1007/978-981-10-0983-9_32>.
- SILVA, P. **Big Data e Data Science: Admirável Mundo Novo**. 2015.