

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE ADMINISTRAÇÃO, CIÊNCIAS CONTÁBEIS, ENGENHARIA DE
PRODUÇÃO E SERVIÇO SOCIAL

LARA MANFRIN PRADO

**DESENVOLVIMENTO DE UMA INTERFACE DE *BUSINESS INTELLIGENCE*
PARA ANÁLISE DE DADOS DE AFASTAMENTOS MÉDICOS EM UMA
MINERADORA**

ITUIUTABA

2021

LARA MANFRIN PRADO

**DESENVOLVIMENTO DE UMA INTERFACE DE *BUSINESS INTELLIGENCE*
PARA ANÁLISE DE DADOS DE AFASTAMENTOS MÉDICOS EM UMA
MINERADORA**

Trabalho de Conclusão de Curso da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de bacharel em Engenharia de Produção.

Orientador: Prof. Dr. Eugênio Pacceli Costa

ITUIUTABA

2021

LARA MANFRIN PRADO

**DESENVOLVIMENTO DE UMA INTERFACE DE *BUSINESS INTELLIGENCE*
PARA ANÁLISE DE DADOS DE AFASTAMENTOS MÉDICOS EM UMA
MINERADORA**

Trabalho de Conclusão de Curso da Universidade Federal de Uberlândia como requisito parcial para obtenção do título de bacharel em Engenharia de Produção.

Ituiutaba, 14 de junho de 2021.

Banca Examinadora:

Prof. Dr. Eugênio Pacceli Costa (orientador), UFU

Prof. Dr. Luís Fernando Magnanini de Almeida, UFU

Prof. Dr. Ricardo Batista Pentead, UFU

*Dedico todo esse trabalho ao meu pai, Nelson, que não está mais presente no plano material,
mas que foi o engenheiro que me inspirou a seguir esse caminho.*

AGRADECIMENTOS

A minha mãe, Maura, verdadeiramente a maior mestre da minha vida que sempre acreditou em mim e que sempre se dispôs a fazer o impossível para me ver bem.

Ao meu pai, Nelson, a minha maior inspiração e orgulho, definitivamente a melhor pessoa que eu já conheci e por quem eu daria de tudo para estar comigo neste momento.

Ao meu irmão Daniel, que sempre esteve do meu lado e é um exemplo para mim de caráter e bondade.

A minha vó, Maria Helena, um anjo na minha vida, que nunca mediu esforços para me ver feliz.

Ao meu namorado, Marcos, que sempre torceu por mim e é meu maior parceiro nessa vida.

A todos os docentes do curso de Engenharia de Produção, pelo apoio nesses anos e por sempre estarem presentes.

Aos discentes do curso da décima segunda turma, que com o passar do tempo se tornaram essenciais nessa jornada: Bruna, Daniel, Fernanda, Gabriela, Henrique, Larissa, Tulio, Vinicius e Vitor.

A todos os meus amigos, que sempre estiveram torcendo por mim.

A todas as outras pessoas que direta ou indiretamente colaboraram com o sucesso desse trabalho.

“A persistência é o caminho do êxito”

Charles Chaplin

RESUMO

Os avanços na área da tecnologia da informação a transformaram em uma grande ferramenta para coleta, análise e elaboração de padrões dos mais diferentes tipos de dados de diferentes empresas. Esse fato propicia oportunidades para o desenvolvimento de aplicações que avaliem dados e forneçam informações relevantes neles contidas. Este trabalho expõe a aplicação de *Business Intelligence* na análise de dados de afastamentos médicos de uma mineradora multinacional. Foram realizados o tratamento dos dados utilizando linguagem de programação em Python e a exibição dos resultados por meio do *software* de visualização de dados, Power Bi. O objetivo desse trabalho foi construir uma plataforma de *Business Intelligence* para disponibilizar informações que poderão ser utilizadas em políticas empresariais visando o bem-estar dos funcionários e, desse modo, melhorando produtividade da empresa, que é demanda premente no mercado atual. Foi desenvolvido um relatório composto por diversos tipos de visualizações, entre gráficos e tabelas, que permite, de maneira dinâmica e clara, a extração de conhecimento.

Palavras-chave: *Business Intelligence*, Python, Power Bi, Afastamentos médicos Mineradora.

ABSTRACT

Advances in the area information technology, have turned it into a great tool for the collecting and analyzing and developing standards for the most different types of data from different companies. This fact provides opportunities for the development of applications that evaluate data and provide relevant information contained therein. This work exposes the application of Business Intelligence in the analysis of medical leave data from a national mining company. The data were processed using Python programming and the results were displayed using the data visualization software, Power Bi. The objective of this work is to create a Business Intelligence platform, to provide information that can be used in corporate policies aimed at the well-being of employees and, thus, improving productivity of a company, which is an urgent demand in the current market. A report was developed, consisting of different types of views, including graphs and tables, which allows, in a dynamic and clear way, to extract knowledge.

Keywords: Business Intelligence, Python, Power Bi, Medical leave, Mining.

LISTA DE FIGURAS

Figura 1 - Dados X Informação X Conhecimento	15
Figura 2 - Importação dos dados No PowerBi	20
Figura 3 - Interface Power Bi	21
Figura 4 - Conectando os dados a uma planilha do Excel	37
Figura 5 - Histórico de alterações feitas à base de dados	37
Figura 6 - Relação entre as tabelas no Power Bi	38
Figura 7 - Medida do IAM feito em linguagem DAX	39
Figura 8 - Segmentadores de dados	39
Figura 9 - Tabela ‘Grupo X Atestados’ exibe o grupo de doença, a quantidade de atestados, a % da quantidade de atestados em relação do total e a somatória de dias de afastamento	44
Figura 10 - Mapa coroplético ‘Quantidade de afastamento por estado’ exibe a quantidade de pessoas afastadas por estado, representado pela saturação da cor azul	45
Figura 11 - Página do relatório ‘Visão Geral’ exibida no Power Bi	46
Figura 12 - Página do relatório ‘Detalhamento Dados’ exibida no Power Bi	47

LISTA DE TABELAS

Tabela 1: Classificação da pesquisa	27
Tabela 2: Descrição Atributos conjunto de dados 'Afastamentos'	29
Tabela 3: Descrição Atributos conjunto de dados 'Lista Funcionários'	36
Tabela 4: Relação entre descrição das etapas e o realizado	48

LISTA DE GRÁFICOS

- Gráfico 1** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao código de carreira (eixo X) 30
- Gráfico 2** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação a idade (eixo X) 31
- Gráfico 3** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tempo de empresa (eixo X) 31
- Gráfico 4** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tipo de afastamento (eixo X) 32
- Gráfico 5** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao status (eixo X) 33
- Gráfico 6** - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tipo de contrato (eixo X) 34
- Gráfico 7** - Gráfico de linha 'Índice de Absenteísmo Médico', relaciona o IAM (eixo Y) e mês (eixo X) 40
- Gráfico 8** - Gráfico de linha 'Soma de dias perdidos', relaciona a soma de dias perdidos (eixo Y) com o mês (eixo X) 40
- Gráfico 9** - Gráfico de barras horizontais 'Análise dos afastamentos por idade', relaciona a frequência de idade (eixo Y) pela quantidade de pessoas afastadas (eixo X) 41
- Gráfico 10** - Gráfico de barras horizontais 'Análise dos afastamentos por tempo de empresa', relaciona a frequência de idade (eixo Y) pela quantidade de pessoas afastadas (eixo X) 42
- Gráfico 11** - Gráfico de barras horizontais 'Afastamentos por cargo', relaciona os cargos (eixo Y) pela quantidade de pessoas afastadas (eixo X) 43
- Gráfico 12** - Gráfico de rosca 'Análise dos afastamentos por gênero', exibe a proporção de gênero em relação a quantidade de pessoas afastadas 43

LISTA DE SIGLAS E ABREVIATURAS

TI	Tecnologia da Informação
BI	<i>Business Intelligence</i>
DAX	<i>Data Analysis Expression</i>
IAM	Índice de Absenteísmo Médico
ETL	<i>Extract, transform, load</i>
ID	Identificador

SUMÁRIO

1	INTRODUÇÃO	13
1.1	CONTEXTUALIZAÇÃO.....	13
1.2	OBJETIVO DA PESQUISA.....	14
2	FUNDAMENTAÇÃO TEÓRICA	15
2.1	DADOS X INFORMAÇÃO X CONHECIMENTO X DECISÃO	15
2.2	SISTEMAS DE INFORMAÇÃO.....	16
2.3	BUSINESS INTELLIGENCE	17
2.3.1.	Benefícios e Dificuldades.....	18
2.3.2	Etapas de implementação de um projeto de BI	19
2.3.3	Power Bi como uma ferramenta de Business Intelligence	19
2.4	LINGUAGEM DE PROGRAMAÇÃO EM PYTHON	21
2.4.2	Pacotes e bibliotecas.....	23
2.5	ÍNDICE DE ABSENTEÍSMO MÉDICO	23
2.6	INDUSTRIA MINERADORA BRASILEIRA	24
3	MÉTODOS DE PESQUISA	26
3.1	CARACTERIZAÇÃO DA PESQUISA	26
3.2	OBTENÇÃO E UTILIZAÇÃO DOS DADOS.....	27
3.3	TRATAMENTO DOS DADOS UTILIZANDO O CONJUNTO ‘AFASTAMENTOS’	28
3.4	TRATAMENTO DOS DADOS UTILIZANDO O CONJUNTO ‘LISTA FUNCIONÁRIOS’	36
3.5	DESENVOLVIMENTO DO RELATÓRIO DE BI.....	36
3.5.1	Importação dos dados.....	36
3.5.2	Criação dos relacionamentos.....	37
3.5.3	Criação do Índice de Absenteísmo Médico.....	38
3.5.4	Criação das visualizações.....	39
4	RESULTADOS	46
5	CONSIDERAÇÕES FINAIS	50
5.1	TRABALHOS FUTUROS	51
	REFERÊNCIAS	52

1 INTRODUÇÃO

1.1 Contextualização

No atual cenário de negócios, todos os processos geram dados, estes vêm se tornando um dos bens mais valiosos nas organizações e, por meio do uso da Tecnologia da Informação esses dados podem ser interligados, coletados, armazenados e disseminados (TEÓFILO E FREITAS, 2007).

Segundo Albertin e Albertin (2008), entre as tantas vantagens que resultaram dos avanços observados nas últimas décadas da Tecnologias da Informação (TI), destacam-se a redução de custo obtida pela integração interna de processos e áreas, aumento da produtividade conseguido pela automação de processos e melhoria da qualidade como resultado da utilização da tecnologia nos produtos e serviços.

Observa-se também o aumento crescente de informações geradas e mantidas no mundo todo e, assim, surgem novos desafios, como a capacidade de obter conhecimento a partir destes recursos, utilizando técnicas como a mineração de dados para criar soluções que tragam vantagens competitivas e boas oportunidades de negócios para as empresas que as utilizam (WITTEN e FRANK, 2005).

Portanto, nesse contexto, torna-se primordial para o sucesso de uma organização a velocidade com que seus gestores e funcionários conseguem identificar e extrair informações que sejam relevantes para a execução de suas atividades (DEAN e WEBB, 2011).

A partir desse pressuposto, foi identificada junto a uma empresa uma oportunidade de implantação de um projeto de *Business Intelligence* (BI) para tratar e exibir informações a partir dos dados de afastamentos médicos, para os quais só havia informações disponíveis em planilhas consolidadas, o que dificultava a visualização da informação e a extração de conhecimento.

A pesquisa em questão é um estudo de caso de natureza quantitativa. A empresa estudada é uma empresa mineradora de grande porte que tem sua sede situada na cidade do Rio de Janeiro – RJ, conta com mais de 50 mil colaboradores em todo o mundo.

Tendo em vista o ambiente o qual a empresa está inserida, torna-se necessária a busca por alternativas para auxiliar a empresa na tomada de decisão. Dessa forma, a implementação de um sistema de BI servirá como um acréscimo para a organização obter melhores resultados.

Sendo uma empresa de grande porte, é natural que o volume de informações sobre os funcionários seja muito grande, desse modo, indicadores são necessários para avaliar a performance da empresa e dos empregados ao longo do tempo.

Inicialmente, foi abordada a teoria e história por trás da linguagem de programação em Python e *Business Inteligente*, e quais os *softwares* que foram utilizados. Para o desenvolvimento do trabalho, foi descrito o tratamento dos dados em Python, a exportação desses dados para o Power Bi e a construção do relatório.

1.2 Objetivo da pesquisa

Este trabalho teve como objetivo geral desenvolver uma solução que disponibilizasse informações que poderão ser utilizadas em políticas empresariais visando o bem-estar dos funcionários.

Para atingir o objetivo geral deste estudo, fez-se necessário também atingir os seguintes objetivos específicos:

- Identificar os dados necessários para a análise;
- Tratar os dados utilizando linguagem de programação em Python;
- Exportar os dados para uma plataforma de *Business Intelligence*;
- Criar um relatório que permita ao usuário identificar de forma clara a informação.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Dados X Informação X Conhecimento X Decisão

Segundo Peixe e Maria (2019), dados são caracterizados por serem simples observações sobre o estado do mundo, são facilmente estruturados, quantificados, transferíveis e são a matéria-prima essencial para o desenvolvimento da informação.

Laudon e Laudon (2004, p. 7) conceituam dado como sendo “[...] correntes de fatos brutos que representam eventos que estão ocorrendo nas organizações ou no ambiente físico, antes de terem sido organizados e arranjados de uma forma que as pessoas possam entendê-los e usá-los.”.

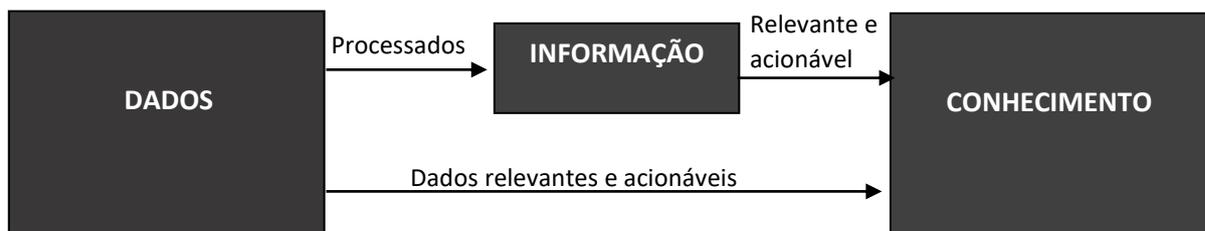
Segundo Peixe e Maria (2019), a informação é o dado configurado de forma pertinente ao entendimento e possibilitando a utilização pelo ser humano.

Conhecimento pode ser entendido com o resultado da interpretação da informação e de sua utilização para alguma finalidade (PIMENTA, 2008).

Nos dizeres de Antonelli (2009), decisão é o processo o qual são escolhidas algumas ou uma alternativa para as ações a serem realizadas. A escolha deve ser baseada no maior conjunto de informações e de conhecimento para que a decisão a ser tomada seja a melhor entre as a disposição.

Segundo Fritzen (2007), as decisões são escolhas feitas com um embasamento substanciado em um propósito. Para que a tomada de decisão ocorra, é necessário fazer escolhas entre diversas alternativas para as ações a serem realizadas e poderem ser tomadas a partir de probabilidades, possibilidades ou alterativas. Turban, McLean e Wetherbe (2004), apresentam o processo de formação do conhecimento como na Figura 1:

Figura 1 - Dados X Informação X Conhecimento



Fonte: Turban, McLean e Wetherbe (2004, p. 327).

A informação no contexto empresarial tem duas finalidades: para conhecimento dos ambientes interno e externo de uma organização e para atuação nestes ambientes.

A análise eficiente de uma grande quantidade de dados pelo homem sem o auxílio de ferramentas é inviável. Dessa forma, torna-se imprescindível o desenvolvimento de ferramentas computacionais que, de forma inteligente e automática, auxiliem o homem a analisar e interpretar esses dados para que se possa desenvolver estratégias de ação em cada contexto de aplicação (GOLDSCHMIDT e PASSOS, 2005).

2.2 Sistemas de Informação

Para O'Brien (2001), um sistema pode ser definido como um conjunto de componentes interligados que formam um todo. Um sistema possui três componentes básicos:

- Entrada: captação de elementos que ingressam no sistema para serem processados;
- Processamento: processos de transformação que convertem insumo(entrada) em produto;
- Saída: transferência de elementos produzidos por um processo de transformação até seu destino.

De acordo com Bazzotti e Elias (2006) o mundo vive a era da informação, exigindo das organizações uma gestão estratégica eficiente. O desenvolvimento e a crescente evolução das organizações são fruto da evolução do conhecimento, os quais podem ser facilitados pela utilização de recursos inteligentes oferecidos pela tecnologia de informação e sistemas de informação.

Segundo Perottoni et. al. (2001), os Sistemas de Informação (SI) provêm da integração de pessoas, tecnologia e organização. Eles auxiliam as empresas a suprirem a necessidade de informações internas e externas, advindas das mudanças que ocorrem no mercado, e sem essas informações as empresas não se tornam aptas a interagir adequadamente no ambiente que se encontram.

Um Sistema de Informação pode ser definido como um conjunto de componentes que coletam, processam, armazenam e distribuem informações com o intuito de contribuir com o planejamento, o controle, a coordenação e o processo decisório em organizações (LAUDON e LAUDON, 2007).

Segundo Reginato e Nascimento (2007), o Sistema de Informação auxilia na interação dos dados entre os usuários para realizar suas atividades e as ferramentas do *Business Intelligence* são usadas no desenvolvimento e administração da informação. Logo, o BI fornece uma visão geral e estratégica do negócio.

2.3 Business Intelligence

Hans Peter Luhn, pesquisador da *International Business Machines* (IBM), usou pela primeira vez o termo *Business Intelligence* em um artigo na década de 1950. O autor designou o termo como a capacidade de compreender as inter-relações dos fatos apresentados de forma concreta para orientar a ação que levará a um objetivo almejado (BOTELHO e FILHO, 2014).

Angeloni e Reis (2006) definem o conceito de *Business Intelligence* como um conjunto de metodologias de gestão, implementadas por meio de ferramentas de *software*, que tem como principal função proporcionar ganhos nos processos decisórios das organizações, baseadas na capacidade analítica das ferramentas que integram em um só lugar todas as informações necessárias ao processo decisório.

Pode se definir o BI como uma ferramenta de inteligência que utiliza dados de várias fontes e os integra para gerar informação e monitoramento do ambiente pertinente para apoiar decisões de negócios por meio de software (PETRINI, POZZEBON e FREITAS, 2004).

O apoio da organização é crítico em projetos de BI, pois estes alteram a forma como os envolvidos acessam a informação (ABUKARI e JOG, 2003). Vale salientar que um projeto de BI pode proporcionar ganhos não somente aos gestores das organizações, mas também a determinados departamentos que precisam se basear em informações coerentes para tomar decisões mais acertadas.

Segundo Turban (2009), os principais objetivos do BI são permitir o acesso interativo aos dados, proporcionar a sua manipulação e fornece a capacidade de realizar a análise adequada da informação aos gerentes e analistas de negócios. Desse modo, o processo de BI baseia-se em uma sequência de passos, iniciando pela transformação de dados em informações, depois em decisões que se tornarão futuras ações.

Seguindo o contexto, Primak (2008, pag. 4) afirma que “[...] o atual interesse pelo BI vem crescendo assustadoramente na medida em que seu emprego possibilita às organizações realizarem uma série de análises e projetos, de forma a agilizar os processos relacionados às tomadas de decisão.”.

Com a alta procura por ferramentas de BI, diversos *softwares* foram desenvolvidos com funcionalidade similares, tais como: conexão a base de dados externas, visualização via tabelas e gráficos, medidas de análises e, em alguns casos, a possibilidade de tratamento dos dados. Contudo, cada *software* possui uma configuração, design e linguagem própria.

2.3.1. Benefícios e Dificuldades

A implementação de um projeto de BI pode trazer grandes benefícios, porém, é a forma de implantação da ferramenta que irá ditar o sucesso ou insucesso da análise.

Segundo Primark (2008), alguns dos benefícios obtidos com um projeto de BI são:

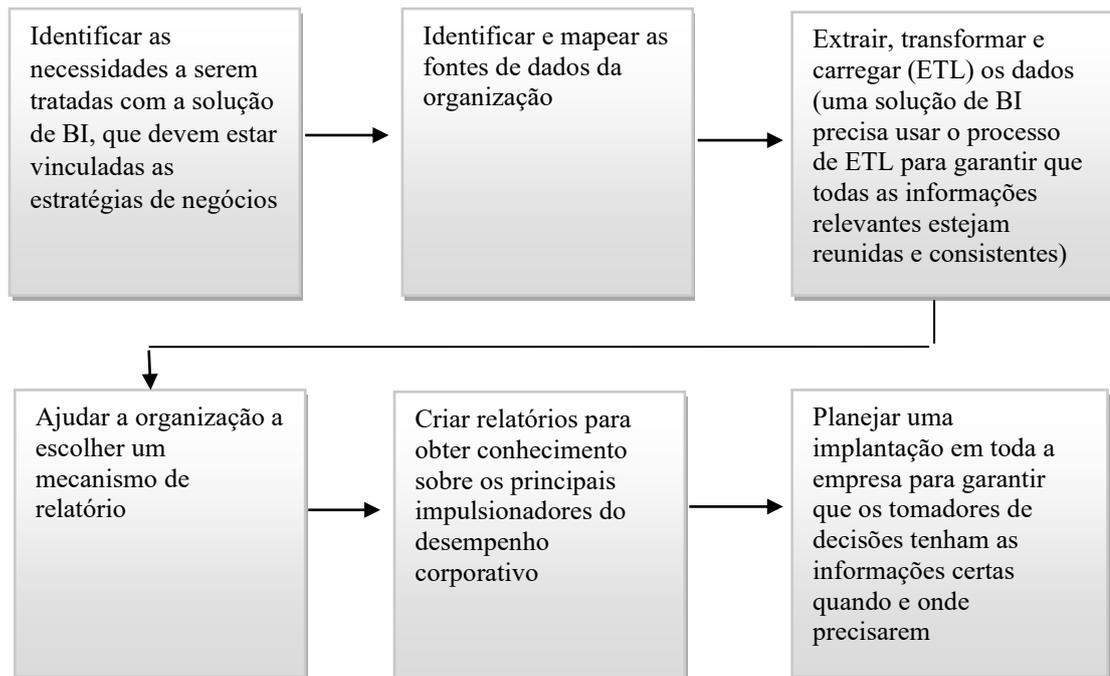
- Redução de custos com *softwares*;
- Redução de custos com administração e suporte;
- Melhor alinhamento dos usuários corporativos;
- Rapidez na informação para a tomada de decisões estratégicas;
- Alinhamento das informações estratégicas e operacionais;
- Informação consistente em vários locais dispersos;
- Vantagem competitiva.

A implementação das ferramentas de BI deve ser bem planejada e Primark (2008) cita as principais dificuldades:

- Ausência de dados úteis para a tomada de decisão causada pela deficiência dos sistemas operacionais;
- Necessidade de uma boa relação entre a área de negócio e a equipe de tecnologia da informação;
- Falta de reconhecimento da necessidade de informação;
- Adoção equivocada de *hardware* e *software*;
- Custo de para implementar um projeto de BI;
- Falta de conhecimentos dos gestores, que pode deixar um projeto de BI sem utilidade na prática;

2.3.2 Etapas de implementação de um projeto de BI

As etapas de implementação de um projeto de BI são importantes para o sucesso de seu uso. Abukari e Jog (2003) sugerem seis passos para uma implementação bem-sucedida, os passos são apresentados no fluxograma abaixo.



2.3.3 Power Bi como uma ferramenta de Business Intelligence

Para que uma informação se torne atraente ao seu receptor, esta precisa ser apresentada de forma a possibilitar rápido entendimento e compreensão.

Segundo Oliveira e Levkowitz (2003), a visualização da informação é parte de um sistema rápido e eficiente na identificação de padrões. Nesse contexto, disponibilizar modelos gráficos e representações visuais possibilita uma interação direta com os usuários para explorar e adquirir conhecimento sobre informações úteis incorporadas nos dados subjacentes.

No presente estudo, o Power Bi ® será utilizado como a ferramenta de visualização de dados, este *software* desenvolvido pela Microsoft é uma ferramenta completa de BI *Service*, com a qual é possível se conectar a diversas fontes de dados e realizar análises.

Com seu lançamento em 2015, o Power BI® surgiu de suplementos que, inicialmente, foram desenvolvidas para o Excel®, o Power Query Editor, Power View e Power Pivot, o que contribuiu para uma rápida aceitação no mercado (LAGO e ALVES, 2018).

Segundo Lago e Alves (2018), outro diferencial do Power BI® é a capacidade de tratamento dos dados, visto que nem todos os concorrentes de mercado têm essa mesma aptidão.

Segundo Cunha e Paula (2019), o Power BI® pode analisar dados de uma grande variedade de fontes, desde uma planilha do Excel®, de um banco de dados e até um serviço de nuvem, sendo compatível também com Facebook® e Google Analytics®, conforme mostra a Figura 2.

Figura 2 - Importação dos dados No PowerBi



Fonte : Vicente, Freitas e Costa (2019).

De acordo com Sá (2020), um projeto em Power BI® pode ser dividido em três etapas. O Power Query Editor é o ambiente no qual ocorre a primeira etapa de um projeto de BI. Nesse ambiente, ocorre a importação dos dados, integração com relatórios de outros sistemas e tratamentos inicial dos dados.

Nesse contexto, segundo Sá (2020), a segunda etapa de um projeto de BI é a modelagem, na qual, são criados relacionamentos entre tabelas, são feitas medidas que processam as linhas da tabela e a manipulação das colunas.

O relacionamento entre tabelas é necessário para exibir as informações e resultados corretamente. Ao importar os dados, se forem duas ou mais tabelas o Power BI® procurará nomes de colunas nas tabelas para determinar se há alguma relação entre elas. Em uma relação, pelo menos uma tabela necessita conter uma lista exclusiva de valores de chave, esse requisito é comum para tecnologias de banco de dados (MICROSOFT, 2020).

Na etapa de modelagem temos uma linguagem chamada DAX (*Data Analysis Expressions*). O DAX auxilia na criação de medidas, que serão novas informações a partir de dados que já estão no modelo (MICROSOFT, 2018).

A última etapa é a da visualização dos dados, o *software* disponibiliza um grande leque de visualizações, tais como: gráficos, tabelas, matrizes, mapas (SÁ, 2020). A partir dos dados importados, podem ser criadas visualizações que exibem diferentes maneiras de se observarem pontos específicos que o usuário necessita no momento da tomada de decisão (VICENTE, FREITAS e COSTA, 2019). A Figura 3 mostra a interface do *software*.

Figura 3 - Interface Power Bi



Fonte: Microsoft (2021).

Conforme apresentado na Figura 4, uma página do Power Bi® pode ser composta por diversas visualizações, tais como gráfico de barras, cartões e gráficos de pizza, todas com um objetivo em comum: transmitir a informação da forma mais clara possível ao receptor.

2.4 Linguagem de programação em Python

Segundo Alves (2016), dentro do campo da tecnologia da informação, a linguagem de programação é a base para qualquer ferramenta da área e pode ter funções desde a criação de um processo automático para preenchimento de planilhas, até funções mais complexas como a criação de um sistema operacional. No presente trabalho, a linguagem de programação utilizada foi em Python, escolhida, principalmente, devido a sua versatilidade e objetividade.

Desde que surgiu em 1991, a linguagem de programação em Python se tornou uma das mais populares entre suas similares, ela se tornou diferenciada por uma numerosa e ativa comunidade de cientistas (MCKINNEY, 2013).

Conforme o relacionamento entre tecnologia e pesquisa aumenta, novas tecnologias que atendam não só ao tratamento de dados numéricos científicos, mas também de alto desempenho e usabilidade, são fundamentais na área científica. O Python é uma linguagem de programação que vem sendo desenvolvida rapidamente e cada vez mais utilizada pela comunidade científica e engenheiros, fornecendo soluções tecnológicas voltadas para as mais diversas áreas e capazes de lidar com esses novos desafios (PÉREZ, GRANGER e HUNTER, 2011).

Sendo uma linguagem interpretada com sintaxe comparada a pseudocódigo, o Python se torna uma linguagem de alto nível adequada para análises científicas. Dentre suas vantagens destacam-se a habilidade de rodar em diversas plataformas, licença de código aberto, sua sintaxe clara e conta com algumas bibliotecas adequadas para atender de forma rápida e precisa a uma gama considerável de cálculos. Essas características tornam essa linguagem bastante eficiente (OLIPHANT, 2007).

Sua curva de aprendizado é alta, sendo um fator importante, pois cientistas costumam escrever programas computacionais não triviais, geralmente utilizando um grande volume de dados, para realizar testes e validar seus experimentos (CATHARINA, 2016).

Segundo Catharina (2016), pode-se citar algumas características da linguagem de programação em Python:

- I. Licença livre de código aberto;
- II. Apresenta um mecanismo de tratamento de erros e exceções;
- III. Possui sintaxe limpa e de fácil interpretação;
- IV. Suporta programação procedural e orientada a objetos;
- V. Por meio de um interpretador, permite programar e testar imediatamente;
- VI. Possui uma forma eficiente de acesso e reutilização de código com o uso de módulos;
- VII. Implementa uma coleta de lixo automático;
- VIII. Apresenta muitos módulos de bibliotecas e pacotes;
- IX. Tem uma famosa comunidade de cientistas que oferecem listas de discussão, dicas, notícias e respostas a dúvidas;

- X. Possui um repositório *online* com uma lista de módulos (bibliotecas) em para atender a diversas finalidades;
- XI. Provê recursos otimizados para atender a longas listas de dados, com gerenciamento de memória e respostas rápidas;
- XII. Suporta listas multidimensionais, efetuando cálculos de forma rápida e precisa;
- XIII. Possui integração com outras linguagens como C, C++, Fortran e Java;
- XIV. Fornece a possibilidade de desenvolver programas de multiusuários e multitarefas com alto desempenho;

2.4.2 Pacotes e bibliotecas

De acordo com Vanderplas (2016), a utilidade do Python para a ciência de dados provém principalmente da grande variedade de pacotes de terceiros, que facilitam o uso da tecnologia para novos negócios e inovação. A seguir são listadas três das principais bibliotecas disponíveis para se utilizar no Python.

- Pandas: a biblioteca Pandas fornece eficientes estruturas de dados e funções que facilitam o trabalho com dados estruturados, tornando o processo mais fácil e rápido. O Pandas é um dos fatores que torna Python um ambiente de análise de dados poderoso e produtivo (MCKINNEY, 2013);
- Numpy: a biblioteca NumPy fornece um objeto de matriz multidimensional para armazenar dados homogêneos ou dados heterogêneos e métodos otimizados para operar neste objeto de matriz (HILPISCH, 2015);
- Matplotlib: esta é a biblioteca de visualização gráfica mais popular para o Python, fornecendo tanto recursos de visualização 2D como 3D (HILPISCH, 2015).

2.5 Índice de Absenteísmo Médico

Segundo Abelha (2021), uma das principais preocupações nas organizações é elaborar um processo de informações que possam contribuir na tomada de decisões. Desse processo, são originadas as medidas de desempenho, que ajudam os administradores a focar pessoas e recursos na direção desejada.

Os indicadores são métricas usadas para descrever a situação de um determinado fenômeno, possibilitando ao usuário fazer comparações, verificar alterações ou tendências e avaliar as ações executadas durante um período (VIEIRA, DETONI e BRAUM, 2006).

Com indicadores bem pensados e acompanhamento, é possível gerenciar desde pessoas e seu desempenho até resultados e a eficácia das estratégias e processos da empresa (SILVA, 2013).

De origem latina, a palavra *absenteísmo* tem como significado "estar fora, afastado ou ausente". No meio corporativo, refere-se à ausência do funcionário no ambiente de trabalho, seja por faltas, saídas ou atrasos, justificadas ou não (PEREIRA, 2019). Ao passo que *absenteísmo* por licença médica pode ser definido como a ausência ao trabalho justificada por meio de licença médica.

Seu efeito é negativo na organização, a ausência dos trabalhadores diminui a produção e reflete diretamente nos indicadores de produtividade e qualidade e, conseqüentemente, na economia (PENATTI, ZAGO e QUELHAS, 2008).

Os índices elevados de *absenteísmo* refletem em todo o clima organizacional, pois revelam aos gestores das organizações uma condição precária ou ineficiente (LACOMBE, 2011).

Segundo Penatti, Zago, Quelhas (2008), cada vez mais o nível de competitividade entre as empresas aumenta, desse modo, faz-se necessário o equilíbrio entre produtividade e capacidade de produção. Assim, o nível de *absenteísmo* deve ser reduzido, bem como a eliminação de suas causas para que, dessa forma, as vantagens competitivas não reduzam.

2.6 Indústria Mineradora Brasileira

A história do Brasil tem íntima relação com a busca e o aproveitamento dos seus recursos minerais, que sempre contribuíram com importantes insumos para a economia nacional. Sendo assim, a mineração torna-se um dos setores básicos da economia do país. (MENDES e VIEIRA, 2010)

Segundo Milioli (1999), o Brasil “destaca-se entre os 10 primeiros na produção de alumínio / bauxita, ouro, manganês, estanho, quartzo, caulim, rocha fosfática, cromo, iemenita, grafite, níquel, terras raras, ligas de ferro, pedras preciosas, amianto, fluorita, magnesita”. A mineração, vista no contexto nacional, envolve a utilização de equipamentos móveis de grande e médio porte, sendo estes, utilizados para diversas tarefas de exploração, beneficiamento e transporte.

De acordo com Souza (2001) na indústria extrativa tem-se a necessidade de se preparar e organizar, nos mais diversos níveis, tanto organizacional, quando operacional, a atuação para exigências em termos de quantidade, qualidade e preço, para garantir a sobrevivência da organização.

Segundo Lima et al (2013), quando os processos produtivos têm como propósito a otimização de recurso e aumentar a produtividade, ao mesmo tempo, torna-se importante manter a segurança e proteção dos trabalhadores para que o trabalho seja realizado de forma correta.

Assim, no contexto da indústria extrativa há condições para incidência de doenças resultantes do trabalho. Conforme Araújo (2012), é um fenômeno complexo e difícil de ser gerenciado, devido às inúmeras situações a que os colaboradores são expostos.

3 MÉTODOS DE PESQUISA

Neste tópico está descrita a metodologia usada em toda a pesquisa, detalhes sobre como foram feitas a obtenção e a utilização dos dados, a seleção de atributos, tratamento dos dados e criação do relatório.

Para o tratamento dos conjuntos de dados, foram utilizadas linguagem de programação em Python e as bibliotecas Pandas e Numpy. Posteriormente ao tratamento do conjunto de dados, eles foram exportados para o *software* Power Bi®, de *Business Intelligence*, quando foram geradas visualizações que deram suporte para a tomada de decisão da empresa. A escolha do *software* Power Bi® foi feita devido ao fato de que ele atende todas as necessidades da empresa, gerando visualizações de dados em tempo real e atualizações automáticas por meio de *dashboards* intuitivos e dinâmicos, facilitando a utilização dos usuários por meio de celulares, tablets e computadores.

Para o presente estudo foram utilizados dois conjuntos de dados: um contendo a informação da identidade (ID) dos empregados e o outro uma compilação de informações sobre os afastamentos médicos de trabalhadores da empresa mineradora, ambos no período de um ano. Foram utilizados dados reais descaracterizados dos empregados.

A empresa em questão é uma mineradora com sua sede situada no estado do Rio de Janeiro e conta com empregados alocados em diversos outros estados do país. O universo a ser investigado compreendeu todos os funcionários em atividade alocados no Brasil, totalizando 43 mil empregados.

3.1 Caracterização da pesquisa

Segundo Gil (2002) existem algumas formas de classificar uma pesquisa, as mais comuns são: quanto a natureza, quanto aos objetivos, quanto aos procedimentos e quanto a forma de abordagem do problema. A Tabela 1 classifica o presente estudo de acordo com as definições citadas.

Tabela 1: Classificação da pesquisa

Definição	Classificação	Justificativa
Natureza	Pesquisa aplicada	Pois objetiva gerar conhecimentos para a solução de problemas específicos.
Objetivos	Pesquisa descritiva	Pois objetiva descrever as características de um fenômeno, estabelecendo uma relação entre as variáveis no objeto de estudo analisado.
Procedimentos	Estudo de caso	Pois consiste no estudo profundo e exaustivo de um ou poucos objetos, de maneira que permita seu amplo e detalhado conhecimento e não é acompanhado de ações.
Forma de Abordagem do problema	Quantitativa	Pois tanto na coleta de informações quanto no tratamento dos dados foram utilizadas técnicas quantitativas.

Fonte: Autoria própria

3.2 Obtenção e Utilização dos dados

Foram obtidos dois conjuntos de dados, o conjunto ‘Afastamentos.xlsx’ com informações relativas aos dados de afastamentos dos empregados, e o conjunto de dados ‘Lista Funcionários.xlsx’ com uma listagem dos identificadores (IDs) dos empregados.

Os dados foram extraídos e disponibilizados pelo setor de Recursos Humanos da empresa, que armazena as informações dos empregados em um sistema de gestão já existente. Dessa forma, problema de garantir a consistência dos dados entre várias fontes, comumente encontrado na área de mineração de dados, não foi enfrentado.

Os dados já passaram por uma pré-avaliação antes de serem inseridos no sistema da empresa, e foi constatado que todos os atributos de ambos os conjuntos de dados estão coerentes a sua descrição e formato, desse modo, garantindo a integridade e consistência informação para desenvolvimento da pesquisa.

Ambos os conjuntos são gerados com periodicidade mensal, o que contribui para a padronização da informação. Optou-se por restringir a pesquisa apenas às informações geradas entre os meses de janeiro de 2018 a dezembro de 2018.

3.3 Tratamento dos dados utilizando o conjunto ‘Afastamentos’

Considerando o objetivo da pesquisa, selecionaram-se quais atributos teriam melhor condição de serem significativos para que fosse possível identificar, em um relatório de BI as principais características de quem esteve afastado. Assim, sendo, os atributos considerados são exibidos na Tabela 2.

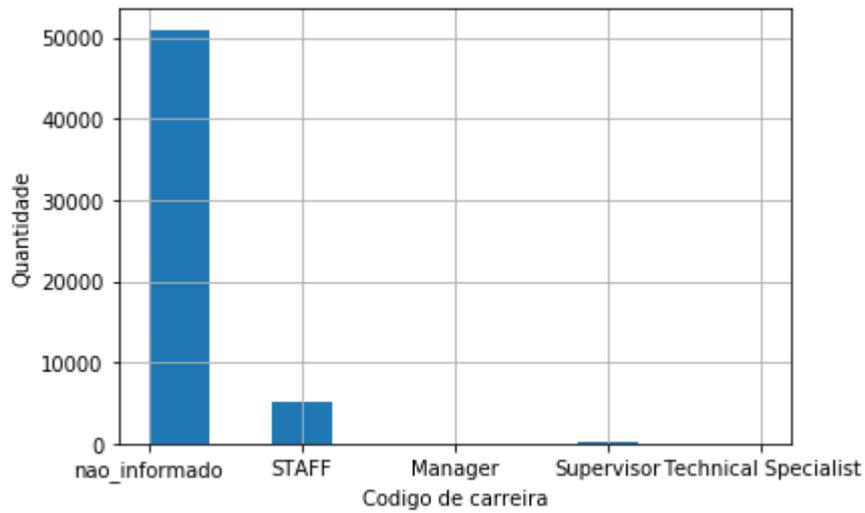
Tabela 2: Descrição Atributos conjunto de dados 'Afastamentos'

Atributo	Descrição
ID	Identificação única de cada funcionário, pode aparecer mais de uma vez no mês, corresponde a quantidade de vezes que o funcionário apresentou um atestado
Mês	Mês de referência
Ano	Ano de referência
Tempo_Empresa	Quantidade de anos que o funcionário está trabalhando na empresa
Idade	Idade do funcionário
Dias_Mês	Quantidade de dias de cada mês no calendário
Status	Classifica o funcionário como ativo ou inativo no mês correspondente, representados, respectivamente pelos caracteres 'A' e 'I'
Gênero	Classificador de gênero
Cargo	Cargo desempenhado pelo funcionário
Codigo_Carreira	Posição do funcionário na hierarquia da empresa
Estado	Estado onde o funcionário trabalha
Contrato	Qual o tipo de contrato do funcionário
Tipo_Afastamento	Classe do afastamento, com 5 classificações distintas: doença não ocupacional, licença maternidade, doença ocupacional, acidente de trabalho típico, acidente de trabalho trajeto.
Dias_Afastamento	Quantidade de dias afastado no mês
Grupo	Grupo de doença que ocasionou o afastamento

Fonte: Autoria própria

A empresa optou por eliminar atributos cuja quantidade de entradas com *missing values* fosse superior a 50%. Como se observa no Gráfico 1, 89% das amostras do atributo 'Cod_Carreira' foram classificadas como 'não informado', desse modo, o atributo foi eliminado.

Gráfico 1 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao código de carreira (eixo X)

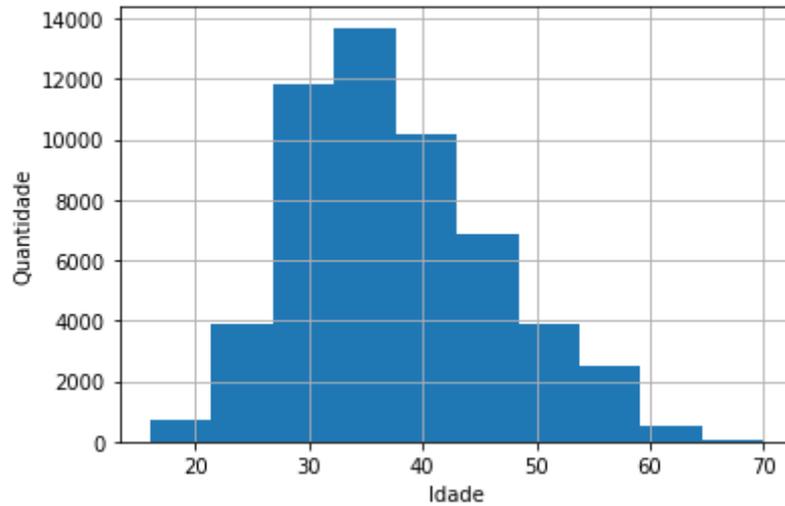


Fonte: A autoria própria.

Os atributos 'Idade' e 'Tempo_Empresa' contém, ambos, 1,1% das entradas nulas. Decidiu-se, junto a empresa, substituir os valores nulos pela média de cada atributo, pois os mesmos serão utilizados para compor outros dois atributos 'Frequencia_Idade' e 'Frequencia_Tempo_Empresa', respectivamente.

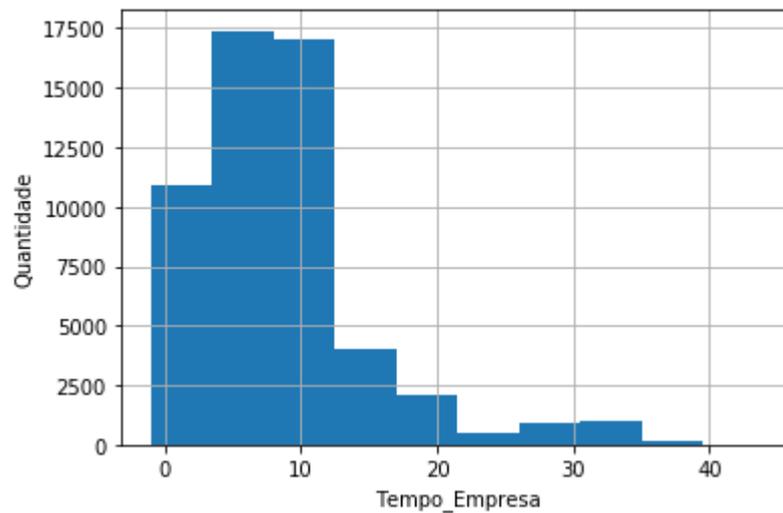
Os atributos 'Idade' e 'Tempo_Empresa' tiveram os valores nulos substituídos por 38 e 6, respectivamente, ambos correspondem ao resultado da média considerando toda a amostra. A Gráfico 2 mostra o gráfico de barra para o atributo 'Idade' e o Gráfico 3 mostra o gráfico de barra para o atributo 'Tempo Empresa'.

Gráfico 2 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação a idade (eixo X)



Fonte: Autoria própria.

Gráfico 3 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tempo de empresa (eixo X)



Fonte: Autoria própria.

Por definição da empresa, os casos em que o tipo de afastamento era devido a Licença Maternidade não foram considerados para o estudo. Os tipos de afastamento considerados e suas definições são listados em sequência:

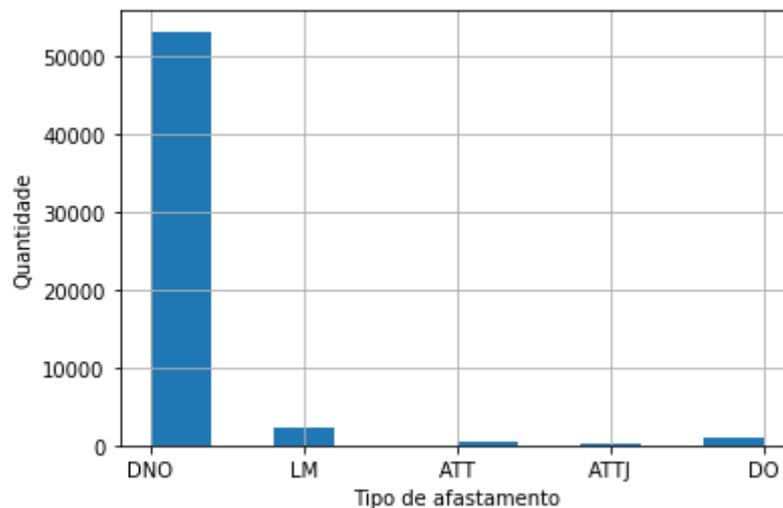
- Doença não Ocupacional: quando não é possível reconhecer nexos entre a enfermidade e a atividade laboral;
- Acidente de Trabalho Típico: acidente que ocorre na execução do trabalho;

- Acidente de Trabalho Trajeto: acidente sofrido pelo trabalhador entre a residência e o local de trabalho, ou vice-versa;
- Doença Ocupacional: enfermidade desencadeada pelo exercício do trabalho.

O Gráfico 4 mostra o gráfico de barras para o atributo ‘Tipo_Afastamento’. Por questões de legibilidade, a legenda no gráfico foi encurtada. Seu significado é apresentado a seguir:

- DNO: Doença não Ocupacional;
- LM: Licença maternidade;
- ATT: Acidente de Trabalho Típico;
- ATTJ: Acidente de Trabalho Trajeto;
- DO: Doença ocupacional.

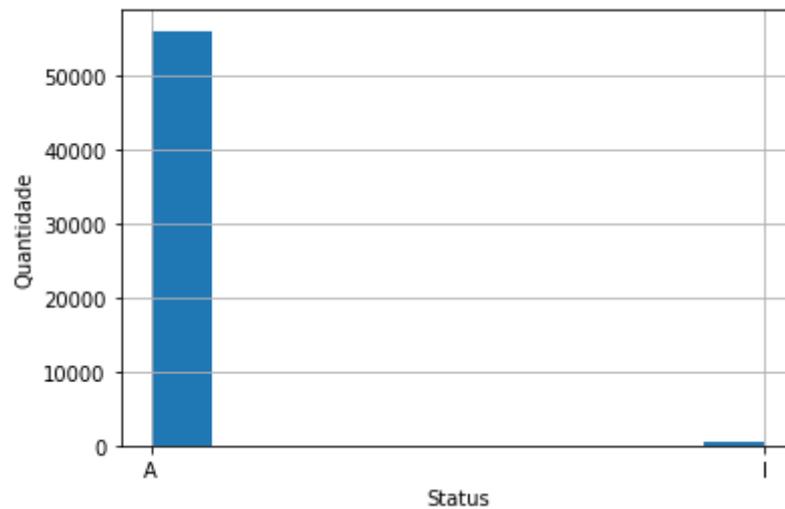
Gráfico 4 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tipo de afastamento (eixo X)



Fonte: Autoria própria.

Além da exclusão de registros relacionados ao tipo de afastamento, observou-se que existiam funcionários inativos, os quais não foram considerados nesse estudo. O Gráfico 5 mostra o gráfico de barras para o atributo ‘Status’.

Gráfico 5 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao status (eixo X)

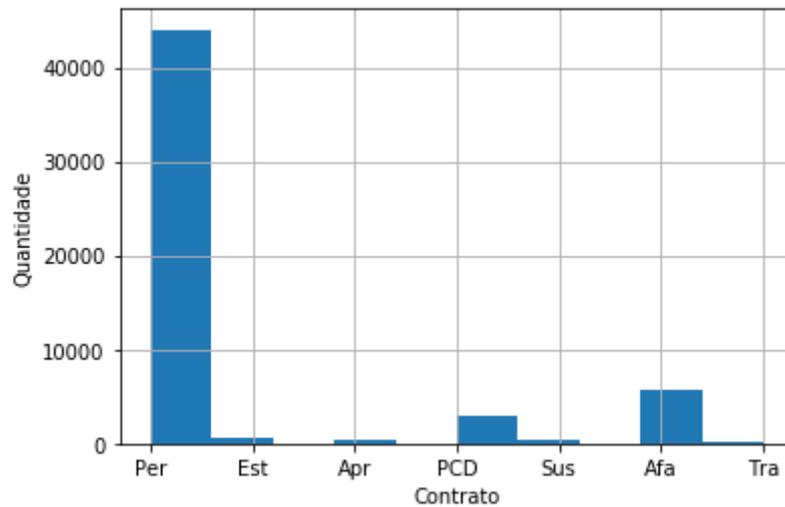


Fonte: Autoria própria.

Por fim, decidiu-se junto a empresa não trabalhar com os casos nos quais o contrato estava suspenso. O Gráfico 6 mostra o gráfico de barras para o atributo 'Contrato'. Por questões de legibilidade, a legenda no gráfico foi encurtada. Seu significado é apresentado a seguir:

- Per: Permanente;
- Est: Estudante;
- Apr: Aprendiz;
- PCD: Pessoa com deficiência;
- Sus: Suspenso;
- Afa: Afastado;
- Tra: Trainee.

Gráfico 6 - Gráfico de Barras da relação da quantidade de indivíduos (eixo Y) em relação ao tipo de contrato (eixo X)



Fonte: Autoria própria.

Após análise individual de cada caso, os *outliers* foram eliminados do espaço amostral. Eliminaram-se 2913 registros do conjunto de dados ‘Afastamentos’, ficando assim com 54114 registros.

Novos atributos foram criados a partir dos atributos ‘Tempo_empresa’ e ‘Idade’, sendo eles: ‘Frequencia_tempo_empresa’ e ‘Frequencia_idade’, respectivamente. Para a criação do atributo ‘Frequencia_tempo_empresa’, o espaço amostral foi dividido em sete categorias, sendo elas:

- 0 a 01 ano;
- 02 a 05 anos;
- 06 a 10 anos;
- 11 a 15 anos;
- 16 a 20 anos;
- 21 a 25 anos;
- 26 ou mais.

Para a criação do atributo ‘Frequencia_tempo_empresa’, o espaço amostral foi dividido em 10 categorias, sendo elas:

- 18 ou menos;

- 19 a 23 anos;
- 24 a 28 anos;
- 29 a 33 anos;
- 34 a 38 anos;
- 39 a 43 anos;
- 44 a 48 anos;
- 49 a 53 anos;
- 54 a 58 anos;
- 59 ou mais.

A divisão das faixas etárias dos atributos 'Frequencia_tempo_empresa' e 'Frequencia_idade' foram feitos seguindo o formato já existente em outros relatórios da empresa, dessa forma, para efeito de comparação adotou -se a mesma.

Foi criado também o atributo 'Chave', este foi utilizado no relatório de BI para relacionar os dois conjuntos de dados utilizados nesse estudo, o atributo em questão é uma concatenação dos atributos 'ID' e 'Mês'.

Os atributos 'Idade', 'Tempo_Empresa', 'Status', 'Contrato' e 'Tipo_Afastamento' não foram necessários para criar as visualizações no relatório de BI, sendo assim, foram eliminados do conjunto de dados. Após o tratamento dos dados, os atributos resultantes estão listados em sequencia.

- ID;
- Mês;
- Ano;
- Gênero;
- Cargo;
- Estado;
- Dias_Afastamento;
- Grupo;
- Frequencia_Tempo_Empresa;
- Frequencia_Idade;
- Chave.

3.4 Tratamento dos dados utilizando o conjunto ‘Lista Funcionários’

Os atributos referentes ao conjunto de dados ‘Lista Funcionários’ são listados na Tabela 3.

Tabela 3: Descrição Atributos conjunto de dados ‘Lista Funcionários’

Atributo	Descrição
ID	Identificação única de cada funcionário
Mês	Mês de Referência

Fonte: Autoria própria

Nesse conjunto, tem-se os IDs de todos os funcionários ativos no período de um ano. Esse conjunto de dados foi usado como suporte para calcular o Índice de Absenteísmo Médico, medida a qual precisa da contagem do total de efetivos no período.

Assim como no conjunto de dados ‘Afastamentos’, foi criado o atributo intitulado ‘Chave’, ele fará a relação entre os dois conjuntos no Power Bi®. Após o tratamento dos dados, os atributos resultantes estão listados em sequencia.

- ID;
- Mês;
- Chave.

3.5 Desenvolvimento do relatório de BI

Nesta etapa, descreveu-se todo o processo de construção do relatório, desde como foi feita a importação dos dados até a etapa de criação das visualizações. O *software* escolhido para construir o relatório foi o Power Bi, por se tratar de uma ferramenta já utilizada pela empresa e por permitir acesso a respostas rápidas por meio da combinação de dados.

3.5.1 Importação dos dados

Os conjuntos de dados resultantes estão no formato Excel (.xlsx). A Figura 4 mostra como é feita a importação da fonte de dados para o Power Bi, a partir de uma máquina local, referenciando um arquivo armazenado em um Excel.

Figura 4 - Conectando os dados a uma planilha do Excel

Obter dados

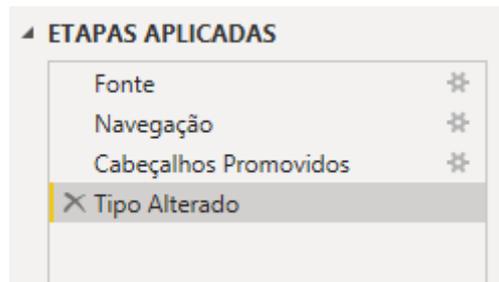
Mais Comum



Fonte: Autoria própria.

A Figura 5 detalha como é registrado o histórico de alterações feitas na base de dados do programa. Essas alterações foram feitas no próprio *software*, sendo elas: correção dos cabeçalhos, alterados os tipos de dados (para: texto, número decimal, data ou número inteiro), entre outras.

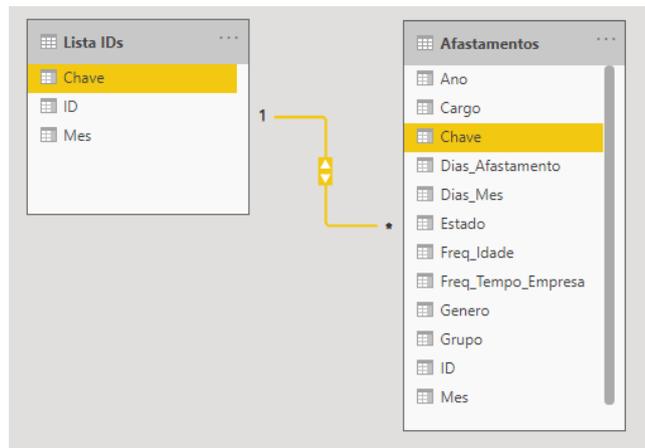
Figura 5 - Histórico de alterações feitas à base de dados



Fonte: Autoria própria.

3.5.2 Criação dos relacionamentos

Para representação do modelo lógico entre as colunas da tabela, utilizou-se a ferramenta de relacionamentos disponível no Power Bi®, o modelo é representado na Figura 6. Nesse modelo, não foi preciso informar quais as colunas a serem relacionadas, pois ele conseguiu identificar automaticamente os relacionamentos.

Figura 6 - Relação entre as tabelas no Power Bi

Fonte: Autoria própria.

A Figura 7 demonstra o relacionamento entre as tabelas ‘Lista IDs’ e ‘Afastamentos’, o campo responsável por relacioná-las é o atributo “Chave”, composto pela matrícula e pelo mês de referência. Na tabela ‘Lista IDs’ os valores do atributo chave são únicos e na tabela “Afastamentos” eles podem aparecer mais de uma vez.

3.5.3 Criação do Índice de Absenteísmo Médico

As medidas são criadas no Power Bi® por meio da linguagem ‘DAX’, que podem ser comparadas ao conceito de fórmulas no Excel®, porém sendo trabalhadas por meio de colunas inteiras. Para realizar algumas análises no Power Bi®, são criadas tais medidas, que são usadas como os valores das visualizações. As medidas ficam salvas caso o usuário deseje aplicar em outra função posteriormente.

O cálculo da medida do IAM é feito a partir de uma fórmula especificada pela empresa e os três fatores citados abaixo fazem parte dela:

- I. A soma dos dias de afastamento (tabela ‘Afastamentos’, atributo ‘Dias_Afastamento’);
- II. A quantidade de dias no mês. Na fórmula é usada a média de dias para caso o usuário queira calcular para mais de um período, porém, se apenas um mês for selecionado, a medida retorna quantidade total de dias do mês selecionado (tabela ‘Afastamentos’, atributo ‘Dias_Mes’);
- III. Contagem do número de IDs (tabela ‘Lista IDs’, atributo ‘ID’).

A Fórmula 1 representa a equação do IAM e a Figura 7 representa a mesma formulação feita em DAX no software.

$$\text{IAM} = \text{Soma}(\text{Dias_Afastamento}) / (\text{Média}(\text{Dias_Mes}) * \text{Contagem}(\text{ID})) \quad (1)$$

Figura 7 - Medida do IAM feito em linguagem DAX

```
% Índice Absenteísmo = DIVIDE(SUM(Afastamentos[Dias_Afastamento]),
(DISTINCTCOUNT('Lista IDs'[ID]) * AVERAGE(Afastamentos[Dias_Mes])) ,
BLANK())
```

Fonte: Autoria própria.

3.5.4 Criação das visualizações

Posteriormente, para a construção do relatório, foram utilizados alguns dos chamados “visuais” disponíveis no Power Bi® que são gráficos, tabelas e filtros personalizáveis refletindo os dados selecionados.

Cada visual foi atrelado às variáveis de interesse, sendo que a variável comum a todos é a quantidade de pessoas afastadas. Na montagem do *layout* foi definida a divisão do relatório em duas páginas, ambas as páginas contêm no canto superior direito três filtros que permitirão ao usuário segmentar os dados por três categorias, sendo elas: mês, estado e cargo. Os filtros são representados na Figura 8.

Figura 8 - Segmentadores de dados



The image shows three dropdown menu filters arranged horizontally. The first filter is labeled 'Mes' and has 'Todos' selected. The second filter is labeled 'Estado' and has 'Todos' selected. The third filter is labeled 'Cargo' and has 'Todos' selected. Each filter has a small downward arrow on the right side of the box.

Fonte: Autoria própria

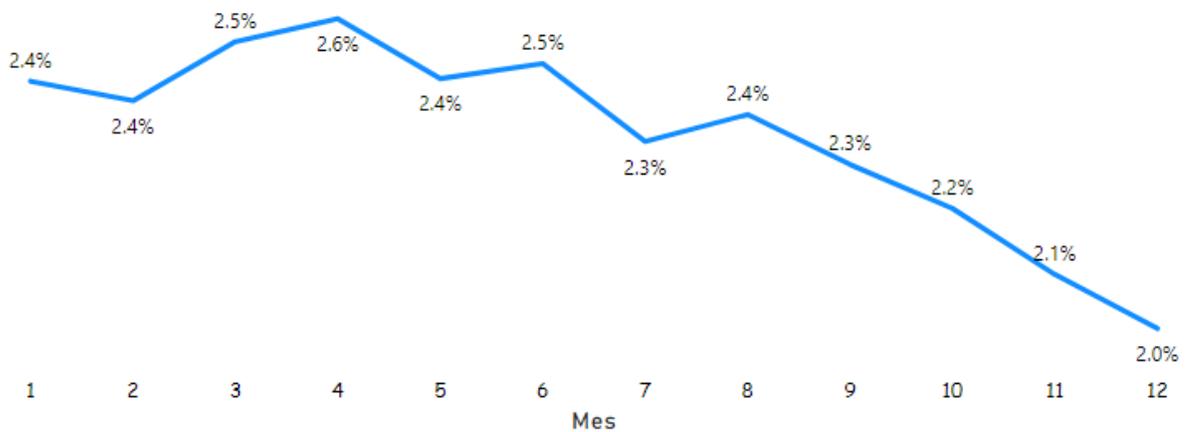
Apresentados na Figura 9, os filtros são ferramentas de relatório que possibilitam ao usuário o refinamento da informação para dar enfoque em alguma variável de interesse. Os filtros podem ter efeito em todas as visualizações da página.

O valor do IAM e a quantidade de dias perdidos são informações complementares e contínuas, que variam na mesma unidade, o mês. Desse modo, escolheu-se exibi-los na

mesma aba e em gráficos que tem como eixo X o atributo mês. A visualização escolhida para apresentar o IAM e a quantidade de dias perdidos é um gráfico de linha. Esse tipo de gráfico permite observar a evolução dos valores e ter uma percepção imediata das tendências além de exibir a informação com clareza. O Gráfico 7 mostra o gráfico ‘Quadro dos Afastamentos’ e o Gráfico 8 mostra o gráfico ‘Soma de dias perdidos’.

Gráfico 7 - Gráfico de linha ‘Índice de Absenteísmo Médico’, relaciona o IAM (eixo Y) e mês (eixo X)

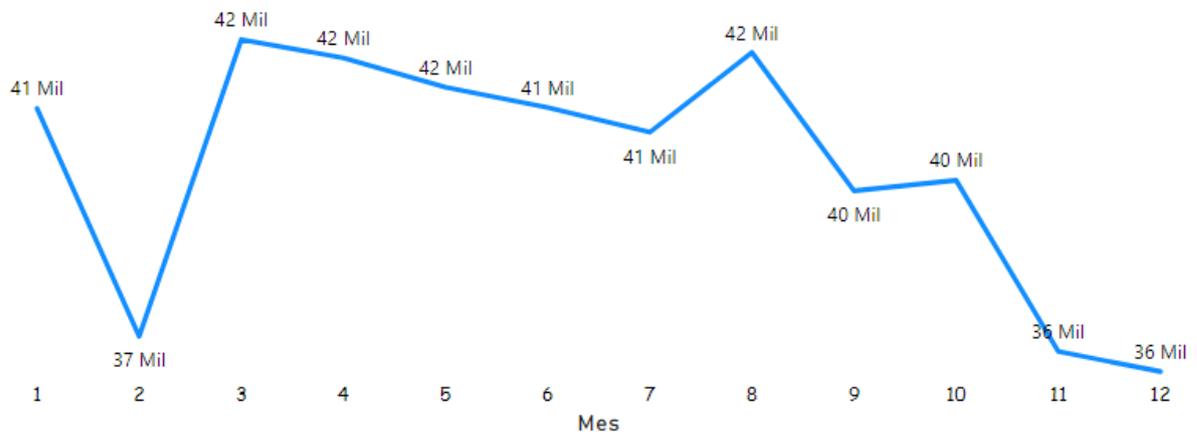
Índice de Absenteísmo Médico



Fonte: Autoria própria.

Gráfico 8 - Gráfico de linha ‘Soma de dias perdidos’, relaciona a soma de dias perdidos (eixo Y) com o mês (eixo X)

Soma de dias perdidos



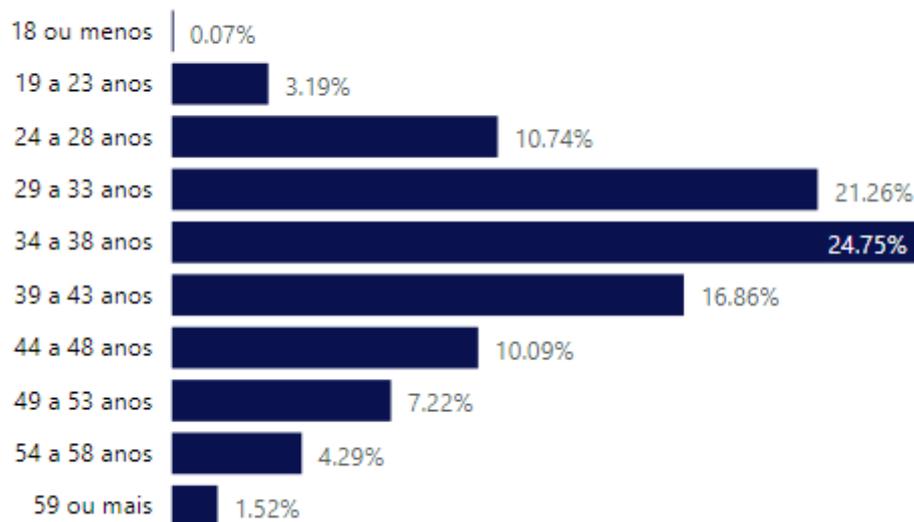
Fonte: Autoria própria.

Para permitir a observação de como os afastamentos se comportam na dimensão de idade e tempo de empresa do funcionário, foram criados os gráficos ‘Quantidade de afastamentos por idade’ e ‘Quantidade de afastamentos por tempo de empresa’, respectivamente.

A visualização escolhida foi o gráfico de barras laterais para ambos, pela sua fácil interpretação. Os atributos ‘Frequencia_Tempo_Empresa’ e ‘Frequencia_Idade’ foram criados especificamente para compor o eixo Y dos dois gráficos, tornando a visualização dos dados segmentados pela unidade de ano mais intuitiva e de fácil leitura. O Gráfico 9 mostra o gráfico ‘Análise dos afastamentos por idade’ e o Gráfico 10 mostra o gráfico ‘Análise dos afastamentos por tempo de empresa’.

Gráfico 9 - Gráfico de barras horizontais ‘Análise dos afastamentos por idade’, relaciona a frequência de idade (eixo Y) pela quantidade de pessoas afastadas (eixo X)

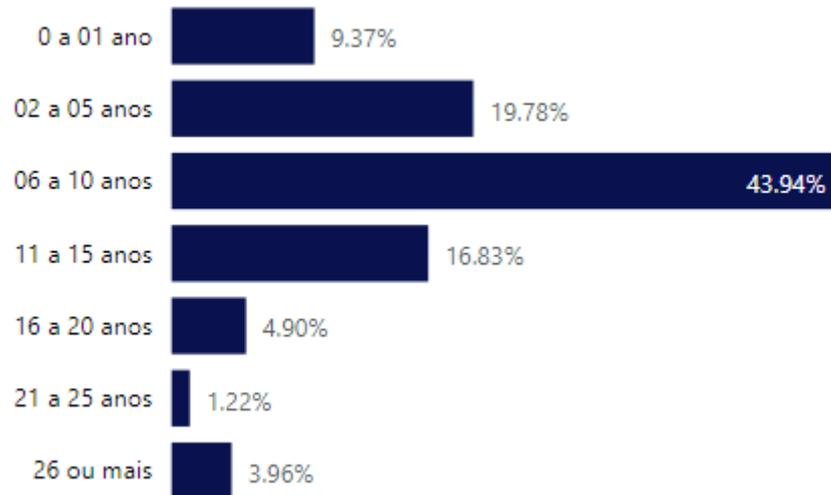
Análise dos afastamentos por idade



Fonte: Autoria própria.

Gráfico 10 - Gráfico de barras horizontais ‘Análise dos afastamentos por tempo de empresa’, relaciona a frequência de idade (eixo Y) pela quantidade de pessoas afastadas (eixo X)

Análise dos afastamentos por tempo de empresa



Fonte: Autoria própria.

O atributo cargo contém 1204 categorias distintas, porém, optou-se por exibir os *top* 85 cargos com a maior quantidade de pessoas afastadas. Os cargos exibidos no gráfico representam 80% de toda a amostra de afastamentos. A visualização escolhida foi o gráfico barras laterais, como mostrado no Gráfico 11.

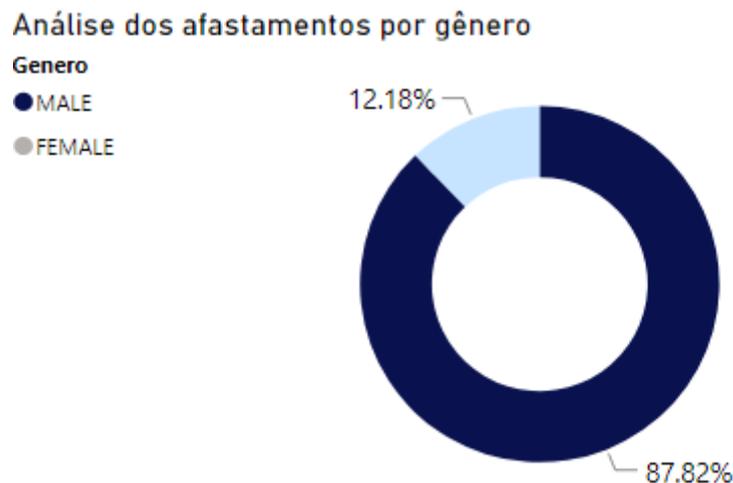
Gráfico 11 - Gráfico de barras horizontais ‘Afastamentos por cargo’, relaciona os cargos (eixo Y) pela quantidade de pessoas afastadas (eixo X)



Fonte: Autoria própria.

Para a informação do atributo gênero, o gráfico escolhido foi o gráfico de rosca, pois, quando se trata de poucas categorias, ele demonstra claramente as diferenças e fornece uma compreensão instantânea das proporções dos dados. O gráfico é exibido no Gráfico 12.

Gráfico 12 - Gráfico de rosca ‘Análise dos afastamentos por gênero’, exibe a proporção de gênero em relação a quantidade de pessoas afastadas



Fonte: Autoria própria.

O atributo Grupo contém a informação que classifica a causa do afastamento do funcionário e contém 22 categorias distintas, optou-se por exibir todas as categorias na página. O tipo de visualização escolhida foi o de tabela pois fornece uma análise mais detalhada dos dados, como mostrado na Figura 9.

Na tabela da Figura 9, retirada do relatório, pode-se analisar cada categoria do grupo de afastamento com sua respectiva quantidade e a porcentagem em relação total de atestados e de dias de afastamento, a mesma encontra-se classificada em ordem decrescente pela coluna ‘Atestados’.

Figura 9 - Tabela ‘Grupo X Atestados’ exibe o grupo de doença, a quantidade de atestados, a % da quantidade de atestados em relação do total e a somatória de dias de afastamento

Grupo	Atestados	% Total Atestados	Dias
Osteomuscular	12883	27.42%	175380
Traumas	5278	11.23%	77415
Respiratório	3960	8.43%	11062
Digestivo	3859	8.21%	19499
Sinais e Sintomas	3642	7.75%	10698
Infecciosas e Parasitárias	3318	7.06%	8843
Oftalmológico	3100	6.60%	15775
Fatores influenciam a saúde	2493	5.31%	11738
Transtornos Mentais e Comportamentais	2093	4.45%	46439
Circulatório	1565	3.33%	16183
Geniturinário	1344	2.86%	8178
Dermatológico	885	1.88%	4712
Neurológico	734	1.56%	7837
Neoplasias (tumores)	599	1.27%	12347

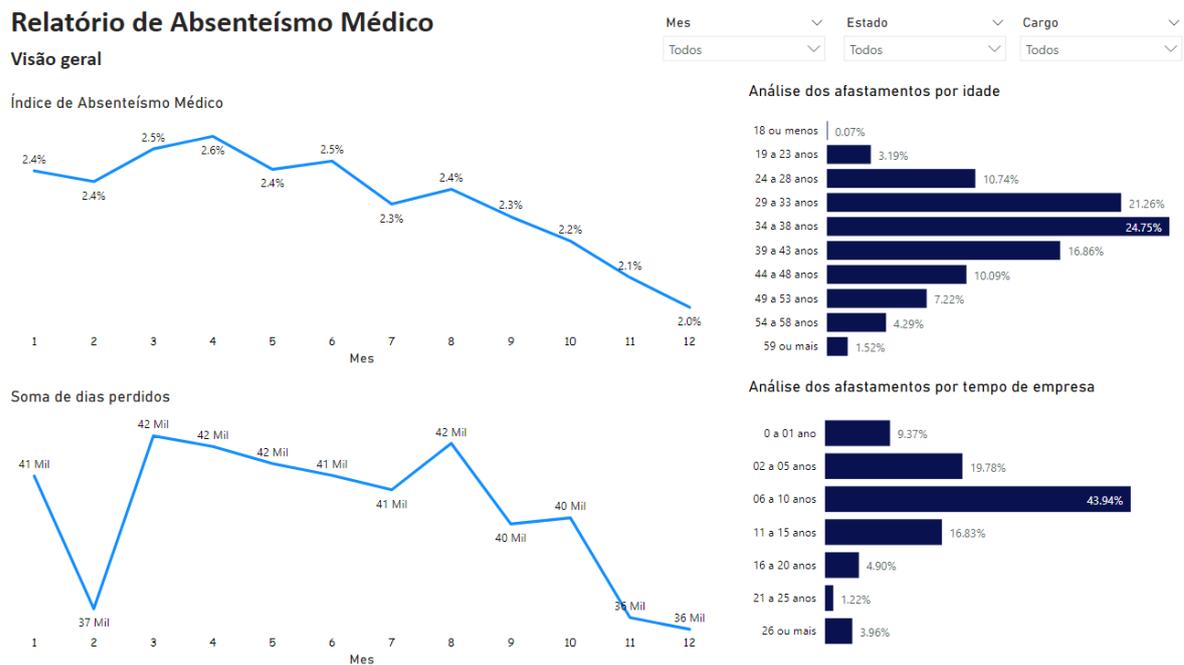
Fonte: Autoria própria.

Com o intuito de demonstrar os dados de maneira mais intuitiva, foi escolhido o gráfico de mapa coroplético para exibir o atributo de localização ‘Estado’. O mapa é exibido

4 RESULTADOS

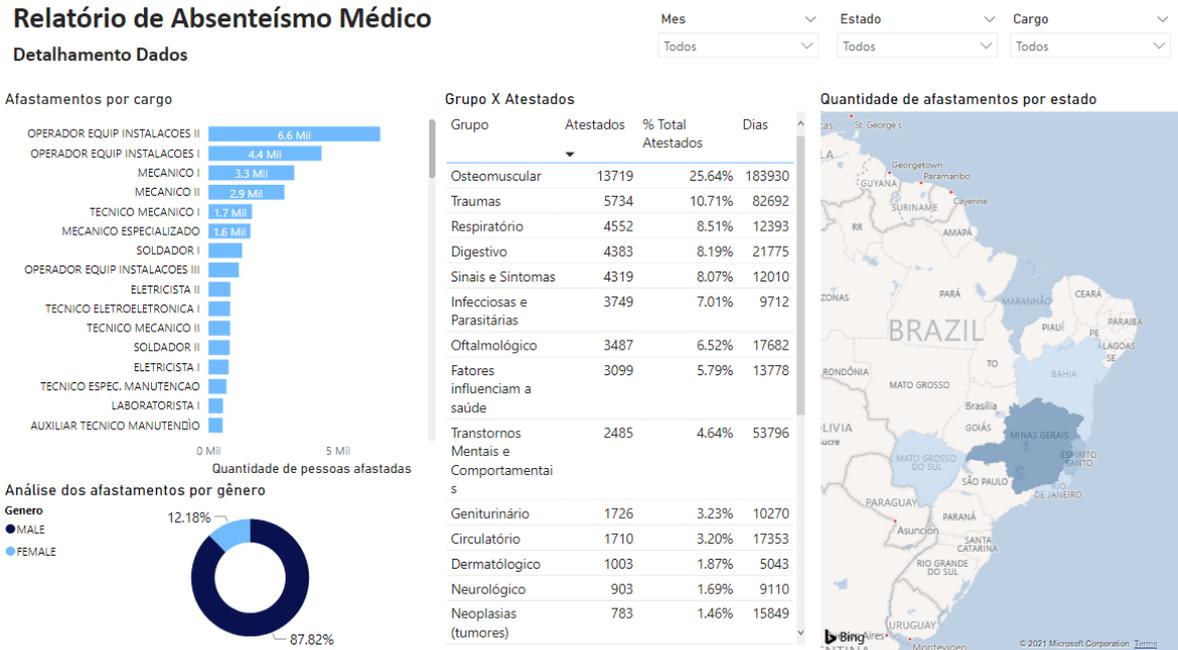
A análise dos dados de afastamentos, representando o período de um ano, de empregados de uma mineradora, tratados em Python e implementados no Power Bi® ficaram disponíveis conforme as Figuras 11 e 12. O relatório ficará exposto em um *workspace* restrito para a gestão e para a equipe médica.

Figura 11 - Página do relatório 'Visão Geral' exibida no Power Bi



Fonte: Autoria própria.

Figura 12 - Página do relatório ‘Detalhamento Dados’ exibida no Power Bi



Fonte: Autoria própria.

Todos os atributos foram utilizados na construção das visualizações. As páginas contêm informações dos afastamentos observadas ao longo dos meses, características gerais dos funcionários afastados e detalhamento dos afastamentos, reunindo informações de cargo, localidade, gênero e grupo de doença. Em ambas as páginas do relatório, os filtros localizados à direita permitem que os usuários segmentem os dados sobre os elementos visuais para dar enfoque nas variáveis de interesse.

A Tabela 4 apresenta a relação das etapas para a construção de um projeto de *Business Intelligence* citadas no Tópico 2.3.2 e que foram realizadas ao longo do estudo.

Tabela 4: Relação entre descrição das etapas e o realizado

Ordem	Descrição das Etapas	Realizado
1	Identificar as necessidades a serem tratados com a solução de BI, elas devem estar vinculadas as estratégias de negócios	Foi identificada junto a empresa a ausência de uma maneira clara de visualizar a informações relacionadas aos afastamentos.
2	Identificar e mapear as fontes de dados da organização	A empresa disponibilizou os dados, que consistiram em dados padronizados e validados no formato de planilhas de Excel.
3	Extrair, transformar e carregar (ETL) os dados.	Os dados foram extraídos e recebidos pela própria empresa e o tratamento dos dados foi feito utilizando linguagem de programação em Python. O resultado do tratamento foram planilhas de Excel que foram carregadas no mecanismo de relatório e continham todos os atributos necessários para criar as visualizações.
4	Ajudar a organização a escolher um mecanismo de relatório	Organização já tinha um mecanismo de relatório definido para projetos de BI, o Power Bi.
5	Criar relatórios para obter conhecimento sobre os principais impulsionadores do desempenho corporativo	Foi criado um relatório que consistiu em agregar todas as informações relevantes sobre os afastamentos da empresa.
6	Planejar uma implantação em toda a empresa para garantir que os tomadores de decisão tenham as informações certas quando e onde precisarem	O relatório ficará publicado em um ambiente corporativo, onde gestores e profissionais da saúde terão acesso, para possam monitorar e identificar informações relevantes.

Fonte: Autoria própria

Com a construção e divulgação do relatório para a gerência, a empresa pode acompanhar de perto e mensurar, com maiores detalhes, os dados dos empregados que foram afastados. Pode se observar que, com a aplicação do tratamento de dados e construção do relatório de BI, as informações de afastamentos ficaram consolidadas em um único local com representações gráficas, intuitivas e dinâmicas.

Ter muitos trabalhadores afastados compromete toda a organização, pois pode gerar sobrecarga de atividades aos que permanecem, impacto na produtividade e prejuízos financeiros.

Desse modo, os responsáveis podem analisar tendências, investigar os fatores e as causas que levam ao afastamento, reduzir a exposição aos riscos e impedir o comprometimento do clima organizacional.

5 CONSIDERAÇÕES FINAIS

O estudo apresentou a importância da implementação de uma ferramenta de BI para o tratamento e na análise de dados. A solução descrita neste trabalho permitiu sintetizar os dados de afastamentos médicos de uma empresa em um relatório, que contribui para o monitoramento mais eficiente da informação. Antes desse trabalho, era possível visualizar esses dados somente no formato de planilhas, que retornam informações agregadas e que visualmente não transmitiam a informação e não permitiam uma análise mais profunda.

O propósito desse trabalho não foi de analisar os dados, e sim, executar a construção de uma ferramenta que desse o suporte para os tomadores de decisão pudessem analisar.

Pode ser visto, de acordo com o que foi abordado nesse trabalho, que com tratamento e a exportação dos dados para o Power Bi®, o usuário pode explorar diferentes formas de visualizações, comparar resultados, analisar pontos de vista, cruzar informações e assim, tirar conclusões que poderão apoiar mais efetivamente tomadas de decisões.

É necessário salientar que apenas a implementação da ferramenta de BI não trará mudanças a empresa. A ferramenta apenas suporta as informações, entretanto, as análises e as tomadas de decisões serão feitas por pessoas que, eventualmente, podem cometer erros. Ainda assim, aplicando o tratamento adequado aos dados, a ferramenta auxilia disponibilizando a informação de maneira intuitiva e consistente, permitindo uma ágil interpretação por parte do usuário.

Durante o desenvolvimento do trabalho foram identificadas dificuldades. Visto que se trata de uma empresa de grande porte, ela já tinha definições e padrões de como cada etapa do processo deveria ser realizada, gerando a uma dependência da empresa para a tomada das decisões e levando a atrasos por necessitar de uma comunicação constante.

As bases de dados aqui utilizadas não atingiram um tamanho que interferisse de forma negativa na implementação das análises. Portanto, o tratamento e a reprodução dos dados nos elementos visuais do relatório apresentaram boa performance.

É importante evidenciar que se trata de uma base de dados acumulativa e, no modo de importação de dados utilizando um arquivo em Excel®, o Power Bi® apresenta uma limitação para o tamanho dele. Desse modo, no futuro, será necessário alterar o tipo de conexão para conexão direta com o banco de dados caso a base chegue no tamanho limite.

Certamente, cada vez mais as organizações estão enxergando o potencial que os projetos de BI têm e fomentando a necessidade de implantá-los. Por outro lado, ainda

dependem de indivíduos com profundo conhecimento do negócio, mas tais análises podem auxiliar enormemente as decisões e interpretação dos fatos.

5.1 Trabalhos futuros

Com os objetivos desse estudo alcançados, oportunidades de melhorias podem ser implementadas. Entre elas estão a mudança da conexão com os dados, passando a ser feita diretamente do banco de dados da empresa que armazena os dados dos empregados, excluindo a necessidade de uma extração manual.

Há a possibilidade da inserção de novos dados com outras informações que poderiam ser relacionadas com a base de afastamentos, como por exemplo, dados sobre a utilização do plano de saúde dos colaboradores.

Sugere-se que seja realizada a construção do relatório utilizando um conjunto maior de dados para analisar a performance do código em Python e do Power Bi®.

REFERÊNCIAS

ABELHA, M. C. **Avaliação dos indicadores de desempenho na gestão de operações de serviços da saúde suplementar**. 104 f. Dissertação (Mestrado) - Curso de Administração e Desenvolvimento Empresarial, Universidade Estácio de Sá., Rio de Janeiro, 2021.

ABUKARI, K.; JOG, V. **Business Intelligence in action**. CMA Management, v. 77, n. 1, p. 15-18, Mar. 2003.

ALBERTIN, L. A.; ALBERTIN, R. M. M. **Benefícios do uso de tecnologia da informação para o desempenho empresarial**. Revista de Administração Pública, p. 275-302, Março/Abril. 2008.

ALVES, W. F. **Estudo Comparativo de Métodos de Regionalização Hidrológica e Desenvolvimento de um Aplicativo para o Gerenciamento de Outorga de Águas Superficiais**. 2016. 97 f. Tese (Doutorado) - Curso de Pós-Graduação em Geociências Aplicadas, Universidade de Brasília, Brasília, 2016.

ANGELONI, M. T.; REIS, Eduardo S. **Business Intelligence como Tecnologia de Suporte a Definição de estratégias para melhoria da qualidade do ensino**. In: Encontro da ANPAD, 2006, Salvador. XXX Encontro Nacional de Pós-Graduação em Administração, 2006. v. 1. p. 16 p.

ARAÚJO, J. P. **Afastamento do trabalho: absenteísmo e presenteísmo em uma Instituição Federal de Ensino Superior**. 2012. Dissertação (Mestrado em Ciências da Saúde) – Universidade de Brasília, Distrito Federal

ANTONELLI, R. A. **Conhecendo o Business Intelligence (BI): uma ferramenta de auxílio à tomada de decisão**. Tecap, Curitiba, v. 3, p. 79-85, 2009. Anual.

BAZZOTTI, C.; ELIAS, G. S. **A importância do sistema de informação gerencial para a tomada de decisões**. Ciências Sociais Aplicadas em revista, Marechal Cândido Rondon, v. 6, n. 11, 2006.

BOTELHO, F. R.; FILHO, E. R. **Conceituando o termo Business Intelligence: origens e principais objetivos**. Sistemas, Cibernética e Informática, Curitiba, v. 11, n. 1, p. 55-60, 2014.

CATHARINA, C. A. B. **Rotina Computacional de Obtenção, Tratamento e Visualização de Dados Oceanográficos**. 2016. 105 f. Tese (Doutorado) - Curso de Instituto de Matemática, Estatística e Física, Universidade Federal do Rio Grande, Rio Grande, 2016.

CUNHA, C.; PAULA, L. B. **Análise do uso de uma ferramenta de Business Intelligence em tomadas de decisão a partir de dados de mídia social**. Revista Científica E-Locução Campinas, n. 16, p. 20-39, maio 2019.

DEAN, D.; WEBB, C. **Recovering from information overload**. McKinsey Quarterly, v. 1, 2011.

FRITZEN, F. **Aspectos comportamentais de tomada de decisão: Um estudo comparativo entre formandos e administração e contabilidade**. 2007. 46 f. Tese (Doutorado) - Curso de Ciências Contábeis, Universidade Federal de Santa Catarina, Florianópolis.

GOLDSCHMIDT, R.; PASSOS, E. **Data Minig: conceitos, técnicas, ferramentas, orientações e aplicações**. Rio de Janeiro: Elsevier Editora Ltda, 2005.

HILPISCH, Y. **Python for Finance**. 1. Ed. United States of America: O'Reilly Media, 2015.

LACOMBE, F. **Recursos Humanos: princípios e tendências**. 2. Ed. São Paulo: Saraiva, 2011.

LAGO, K.; ALVES, L. **Dominando o Power BI**. 1 ed. Karine Gouvea Dibai do Lago, 2018.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de informação gerenciais: administrando a empresa digital**. 5 Ed. São Paulo: Pearson Prentice Hall, 2004.

LAUDON, K. C.; LAUDON, J. P. **Sistemas de Informações Gerenciais**. 7. Ed. São Paulo: Pearson Prentice Hall, 2007.

LIMA, L.; ARAÚJO, N.; SILVA, R. **A relação entre segurança no trabalho e produtividade dos funcionários de um canteiro de obras LIMA em João Pessoa/PB.** Teoria e prática na engenharia civil, São Paulo, n. 22, p. 51-60, 2013

MATTAR, F. Najib. **Pesquisa de Marketing.** Vol 1 e 2, Sao Paulo: Atlas, 2005

MCKINNEY, W. **Python for data Analysis.** Sebastopol, California: O'Riley Midia, 2013.

MENDES, A.; VIEIRA, R. **Gestão Ambiental na Industria de Mineração do Brasil: Técnicas para minimizar potenciais impactos ambientais na extração do cobre.** 2010. 10 f. Tese (Doutorado) - Bauru, 2010.

MICROSOFT. **Criar e gerar relações no Power Bi Desktop.** 2020. Disponível em: <<https://docs.microsoft.com/pt-br/power-bi/transform-model/desktop-create-and-manage-relationships>>. Acesso em 04/02/2021.

MICROSOFT. **Início rápido: Aprenda os fundamentos de DAX em 30 minutos.** 2018. Disponível em: <<https://support.microsoft.com/pt-br/office/in%3%adcio-r%3%a1pido-aprenda-os-fundamentos-de-dax-em-30-minutos-51744643-c2a5-436a-bdf6-c895762bec1a?ui=pt-br&rs=pt-br&ad=br>> Acesso em: 02/12/2020.

MICROSOFT. **Introdução a dashboards para designers do Power Bi.** 2021 Disponível em <<https://docs.microsoft.com/pt-br/power-bi/create-reports/service-dashboards>>. Acesso em 02/12/2020.

MILIOLI, G. **Abordagem ecossistêmica para a mineração: uma perspectiva comparativa para Brasil e Canadá.** Tese (Programa de Pós-Graduação em Engenharia de Produção), Departamento de Engenharia de Produção, Universidade Federal de Santa Catarina, 1999.

O'BRIEN, J. A. **Sistemas de informação e as decisões gerenciais na Era da internet.** 9 ed. São Paulo: Saraiva, 2001

OLIPHANT, T. E. **Python for Scientific Computing.** Computing in Science & Engineeng, v.9, p. 10–20. May – June 2007.

OLIVEIRA, M. C. F.; LEVKOWITZ, H. **From visual data exploration to visual data mining: a survey**. IEEE Transactions on Visualization and Computer Graphics, vol. 9, no. 3, pp. 378-394, July-Sept. 2003.

PEIXE, A.; MARIA, A. **Gestão da Informação: temas e abordagens**. Curitiba. 2019. 53 p.

PENATTI, I.; ZAGO, J.; QUELHAS, O. **Absenteísmo: As consequências na gestão de pessoas**. 2008. 11 f. Tese (Doutorado) - Universidade Federal Fluminense, Niterói, 2008.

PEREIRA, L. B. P. **Absenteísmo por saúde: uma experiência em uma empresa pública mineira**. 2019. 47 f. Tese (Doutorado) - Curso de Especialização em Gestão de Pública, Fundação João Pinheiro, Belo Horizonte, 2019.

PÉREZ F.; GRANGER B. E.; HUNTER.J.D. **Python: An Ecosystem for Scientific Computing**. *Computing in Science & Engineeng*, Berkley, v. 13, p 13-21, 2011.

PEROTTONI, R.; OLIVEIRA, M.; LUCIANO, E.M.; FREITAS, H. **Sistemas de informações: um estudo comparativo das características tradicionais às atuais**. Porto Alegre: ReAd (<http://read.adm.ufrgs.br>), PPGA/EA/UFRGS, v.7, n. 3, 2001

PETRINI, M.; POZZEBON, M.; FREITAS, M. T. **Qual é o papel da inteligência de negócios (BI) nos países em desenvolvimento? Um panorama das empresas brasileiras**. In: Enanpad – Encontro Nacional de Pos-graduação em Administração, XXVIII. 2004, Curitiba.

PIMENTA, R. C. de Q. **Gestão da informação: um estudo de caso em um instituto de pesquisa tecnológica**. Dissertação (Mestrado em Administração) - Universidade Federal do Rio Grande do Norte, Natal,2008.

PRIMAK, F. V. **Decisões com B.I. (Business Intelligence)**. Rio de Janeiro: Ciência Moderna, 2008.

PROENG. **O que é Business Intelligence- Inteligência Empresarial?** .2020 Disponível em: <https://proengct.com.br/o-que-e-business-intelligence-inteligencia-empresarial/>

REGINATO, L.; NASCIMENTO, A. M. **Um estudo de caso envolvendo Business Intelligence como instrumento de apoio à controladoria.** Revista Contabilidade & Finanças, São Paulo, v. 18, n. 1, 2007.

SÁ, M. P. B. **Lei de Benford e Regras de Associação no Power BI: Ferramentas Estatísticas Aplicadas à Auditoria.** 2020. 66 f. Monografia (Especialização) - Curso de Estatística Aplicada, Universidade Federal da Paraíba, João Pessoa, 2020.

SILVA, A. F. **Indicadores de Desempenho: Um estudo de caso na empresa Net Serviço.** 2013. 58 f. Tese (Doutorado) - Curso de E Ciências Sociais, Universidade Federal da Paraíba, João Pessoa, 2013.

SOUZA, J. C. **A manutenção produtiva total na indústria extrativa mineral: a metodologia TPM como suporte de mudanças.** 150 f. Dissertação (Mestrado) - Curso de Engenharia de Produção, Universidade Federal de Santa Catarina, Florianópolis, 2001.

TEÓFILO, R. B.; FREITAS, L. S. **O uso de tecnologia da informação como ferramenta de gestão.** In: Simpódio de Excelência em Gestão e Tecnologia, IV, 2007, Rio de Janeiro.

TURBAN, E. **Business Intelligence: um enfoque gerencial para a inteligência do negócio.** São Paulo: Bookman, 2009.

TURBAN, E.; MCLEAN, E.; WETHERBE, J. **Tecnologia da Informação para gestão: transformando os negócios na economia digital.** 3 ed. Porto Alegre: Bookman, 2004.

VANDERPLAS, J. **Python Data Science Handbook.** Sebastopol: O'Reilly, 2016.

VICENTE, A.M.; FREITAS, G.C.; COSTA, J.C.L. **Gerenciamento e análise de dados da gestão da cadeia de suprimentos aplicando as ferramentas CIS-ERP e Power Bi.** 2019. 12 f. Tese (Doutorado) - Faculdade de Tecnologia de São José dos Campos, São José dos Campos, 2019.

VIEIRA, D. K.; DETONI, J. D.; BRAUM, L. M. **Indicadores de Qualidade em uma Unidade Hospitalar**. In: Simpósio de Excelência em Gestão e Tecnologia, III., 2006, Curitiba.

WITTEN, I H.; FRANK, E. **Data Mining: practical machine learning tools and techniques**. 2a Edição. San Francisco: Editora Morgan Kaufmann, 2005.