FireData
Everymind

Controle do Documento

Histórico de revisões

Data	Autor	Versão	Resumo da atividade
09/08/2022	João Alcaraz	1.0	Criação do documento
09/08/2022	Alexandre Fonseca, Gabriela Rodriguês João Alcaraz, Filipi Kikuchi, Lucas Pereira e Sofia Pimazzoni	1.1	Introdução, Análise SWOT e Value Proposition Canvas do produto
09/08/2022	Alexandre Fonseca, Gabriela Rodriguês João Alcaraz, Filipi Kikuchi, Lucas Pereira e Sofia Pimazzoni	1.2	Contexto da indústria e Matriz de Riscos
11/08/2022	Bruno Meira, João Alcaraz e Filipi Kikuchi	2.0	Descrição dos dados
11/08/2022	Gabriela Rodrigues, Lucas Pereira e Sofia Pimazzoni	2.1	Primeira versão dos gráficos da relação entre as variáveis do gráfico.
12/08/2022	Bruno Meira,, Gabriela Rodriguês João Alcaraz, Filipi Kikuchi, Lucas Pereira e Sofia Pimazzoni	2.2	Objetivos e Justificativa, Descrição dos dados a serem utilizados, Descrição estatística básica dos dados e Descrição da predição desejada.

Sumário

1. Introdução

A Everymind é uma empresa de consultoria especializada na aplicação da Salesforce, utilizando diversas ferramentas e tecnologias para fornecer soluções personalizadas. Por ser uma das maiores na América Latina e integrar o grupo Uol, possui uma grande área de atuação, com presença nacional (15 estados brasileiros em 5 regiões) e internacional (escritórios no Japão e Europa).

O problema expõe a preocupação da empresa com sua taxa de Turnover e na descoberta dos fatores que mais contribuem para a rotatividade de funcionários. Dessa forma, a empresa propõe a modelagem de um modelo preditivo que indique possíveis tendências de saída.

2. Objetivos e Justificativa

2.1. Objetivos

Descreva resumidamente os objetivos gerais e específicos do seu parceiro de negócios

A Everymind, como uma empresa que aplica o modelo Salesforce, possui o objetivo de gerenciar projetos dos seus clientes, a fim de aumentar a performance do negócio com excelência e qualidade. Através das reuniões e encontros com o parceiro, pudemos inferir alguns pontos relevantes. Devido à expansão da empresa no recente contexto pandêmico, ela conquistou seu espaço como um dos principais players no segmento. Dessa forma, manter a área de seus serviços expandida é de grande importância para o parceiro. Especificamente, a taxa de rotatividade da empresa é alta e entender os fatores que contribuem para a saída dos funcionários é um objeto de desejo interno da instituição.

2.2. Justificativa

Faça uma breve defesa de sua proposta de solução, escreva sobre seus potenciais, seus benefícios e como ela se diferencia.

A solução proposta tem o potencial de ajudar a Everymind com seu maior problema, o Turnover, com um modelo preditivo criado especificamente para a base de colaboradores da empresa, que irá prever quais funcionários têm tendência a sair.

3. Metodologia

Descreva as etapas metodológicas que foram utilizadas para o desenvolvimento, citando o referencial teórico. Você deve apenas enunciar os métodos, sem dizer ainda como ele foi aplicado e quais resultados obtidos.

3.1. CRISP-DM

Descreva brevemente a metodologia CRISP-DM e suas etapas de processo

3.2. Ferramentas

Descreva brevemente as ferramentas utilizadas e seus papéis (Google Colaboratory)

3.3. Principais técnicas empregadas

Descreva brevemente as principais técnicas empregadas, algoritmos e seus benefícios

4. Desenvolvimento e Resultados

De maneira geral, você deve descrever nesta seção a aplicação dos métodos aprendidos e os resultados obtidos por seu grupo em seu projeto

4.1. Compreensão do Problema

4.1.1. Contexto da indústria

O contexto da indústria pode ser definido pelas 5 forças de Porter, sendo elas rivalidade entre os concorrentes atuais, poder de barganha dos fornecedores, poder de barganha dos clientes, ameaça de novos entrantes, ameaça de novos produtos ou serviços substitutos.

Abaixo está uma pequena introdução e logo em seguida as 5 forças de Porter da empresa.

Players do mercado: Dentre os principais players do mercado, estão as empresas: Imaginedone, SYS4B, JFox. São empresas de consultoria que fazem uso da Salesforce e estão, assim como a Everymind, em ascensão neste ramo da indústria.

O mercado de trabalho pós pandemia alterou o ambiente empresarial. O trabalho que antes era feito de forma presencial agora é feito de forma híbrida (em sua maioria) e homeoffice; com isso a comunicação entre funcionários é uma das principais preocupações visto que muitos dos colaboradores nem mesmo vivem no mesmo estado que a empresa. Além disso, o uso das nuvens para alocar dados é uma das principais tendências.

Modelo de negócio: O objetivo é prever quais funcionários vão ficar e quais vão sair da empresa, analisando os dados coletados pelos stakeholders. Ao fazer a análise dos motivos causadores da alta rotatividade na empresa, o projeto espera melhorar o turnover de funcionários, além de atrair e manter mais funcionários na Everymind. Os recursos utilizados serão o Google Colab, o Python e o Pandas.

RIVALIDADE ENTRE OS CONCORRENTES

- Muitas empresas que prestam o mesmo serviço (Imaginedone, SYS4B,JFox,etc)

PODER DE BARGANHA ENTRE OS FORNECEDORES

- Migração dos fornecedores para ofertas mais vantajosas
- Poucos profissionais especializados em Salesforce

PODER DE BARGANHA DOS CLIENTES

- Clientes conseguem negociar e personalizar serviços com outros fornecedores
- Consulta fácil à outros serviços, permitindo comparações ágeis
- Cliente não é fiel a marca

AMEAÇA DE NOVOS ENTRANTES

- Baixa barreira para empresas existentes entrar nesse mercado
- Alta demanda pelo serviço gera propostas melhores e diferentes

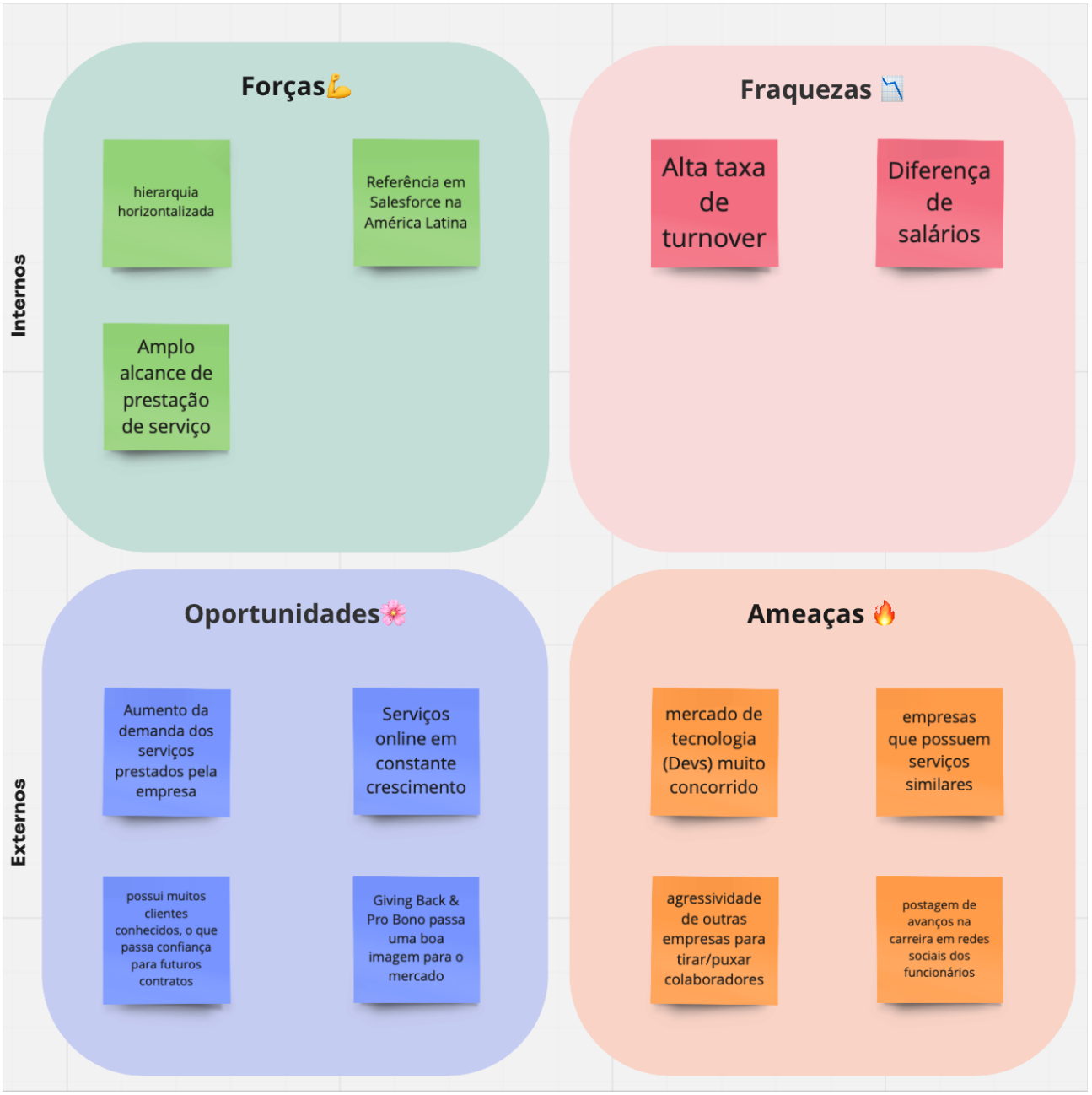
AMEAÇA DE NOVOS PRODUTOS OU SERVIÇOS SUBSTITUTOS

- Serviços com maior nível de personalização e eficiência
- Ofertas mais vantajosas

4.1.2. Análise SWOT

A análise SWOT é uma ferramenta que utiliza de quatro aspectos que ajudam a visualizar a posição de certa empresa no mercado. Os 4 aspectos são Forças, Fraquezas, Oportunidades e Ameaças, e são divididos entre, internos e externos, ou seja, se a empresa tem influência sobre tal fator do aspecto ou não.

Abaixo está a análise SWOT realizada pelo grupo:



4.1.3. Planejamento Geral da Solução

A descrição da solução é uma ferramenta de entendimento de negócio utilizada para esclarecer e compreender a melhor maneira qual a solução e como ela será abordada durante o projeto. A nossa solução segue o pequeno roteiro abaixo:

a) quais os dados disponíveis (fonte e conteúdo - exemplo: dados da área de Compras da empresa descrevendo seus fornecedores)

b) qual a solução proposta (pode ser um resumo do texto da seção 2.2)

c) qual o tipo de tarefa (regressão ou classificação)

d) como a solução proposta deverá ser utilizada

e) quais os benefícios trazidos pela solução proposta

f) qual será o critério de sucesso e qual medida será utilizada para o avaliar

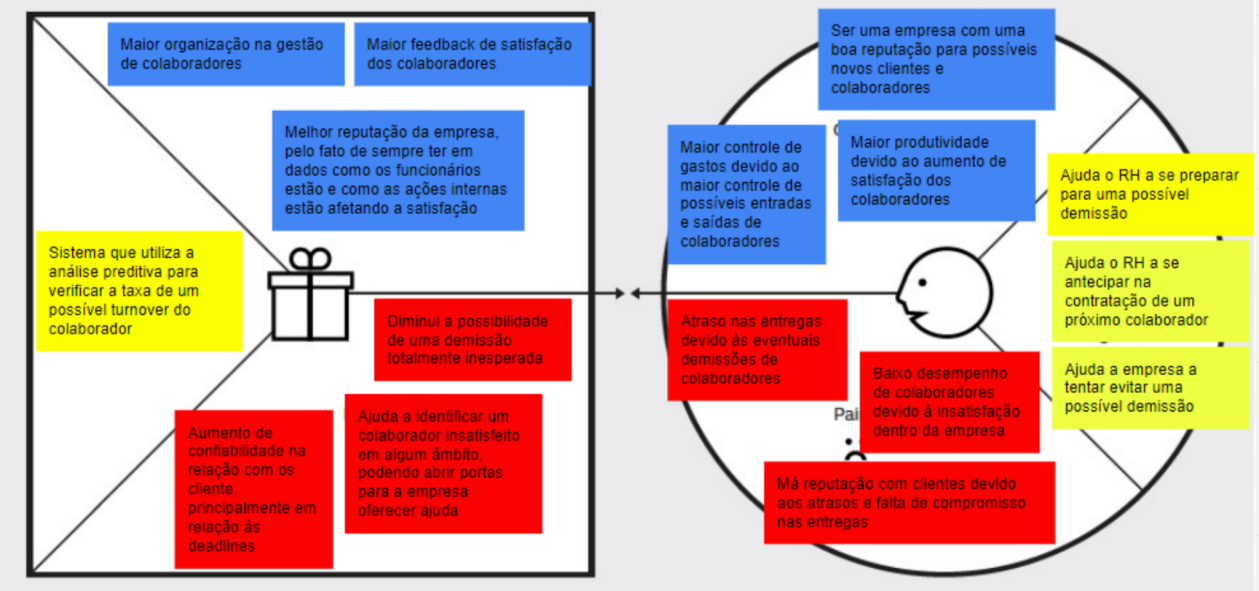
Abaixo é possível visualizar a descrição da solução proposta pelo nosso grupo:

A Everymind é uma empresa certificada SalesForce que está preocupada com sua alta taxa de Turnover e possui dificuldades em analisar os motivos responsáveis pela alta rotatividade. Neste contexto, a empresa forneceu dados referentes aos colaboradores (salário, cargo, data de entrada/saída, etc.) para que possamos desenvolver um modelo preditivo de classificação capaz de identificar quais funcionários têm tendências a sair ou permanecer no trabalho. A solução proposta deverá ser usada como complementação do sistema atual pelo setor de RH, facilitando o processo, e como uma ferramenta de análise que possibilite a identificação de eventuais fatores que contribuem para uma demissão a fim de diagnosticá-los e contribuir para a tomada da melhor decisão possível. Consequentemente, isso será benéfico para a Everymind pois é melhor para uma empresa manter os funcionários a longo prazo. Dessa forma, serão usados dois critérios: a taxa de precisão do algoritmo e a taxa de Turnover da empresa para avaliar o desempenho da solução.

4.1.4. Value Proposition Canvas

A proposta de valor é uma ferramenta importantíssima de negócio, pois ela permite que a empresa visualize se o produto que deseja desenvolver está em uma boa posição no mercado, por exemplo, se ele resolve as dores do público alvo e o que o seu produto tem de diferencial comparado a outros já existentes.

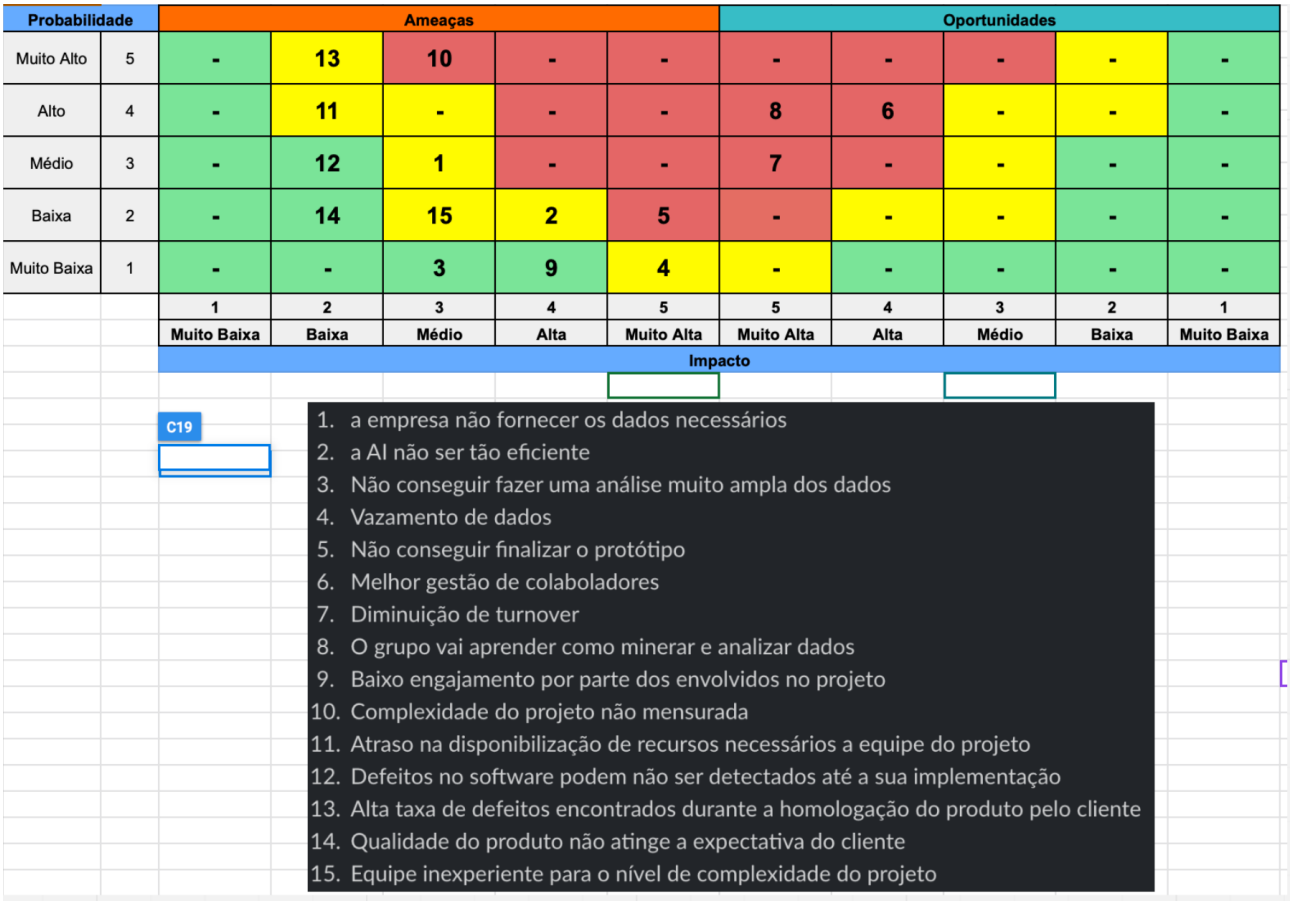
Logo abaixo está a Proposta de valor desenvolvida pelo grupo:



4.1.5. Matriz de Riscos

A última ferramenta importante no entendimento de negócios é a matriz de risco. Sua função é ajudar a empresa a tomar decisões baseadas nos impactos e na probabilidade de certos riscos acontecerem, tanto com o projeto como com a empresa, sendo eles oportunidades ou não.

Abaixo é possível visualizar a matriz de risco desenvolvida pelo grupo:



Probabilidade		Ameaças					Oportunidades				
Muito Alto	5	-	13	10	-	-	-	-	-	-	-
Alto	4	-	11	-	-	-	8	6	-	-	-
Médio	3	-	12	1	-	-	7	-	-	-	-
Baixa	2	-	14	15	2	5	-	-	-	-	-
Muito Baixa	1	-	-	3	9	4	-	-	-	-	-
		1	2	3	4	5	5	4	3	2	1
		Muito Baixa	Baixa	Médio	Alta	Muito Alta	Muito Alta	Alta	Médio	Baixa	Muito Baixa
		Impacto									

C19

1. a empresa não fornecer os dados necessários
2. a AI não ser tão eficiente
3. Não conseguir fazer uma análise muito ampla dos dados
4. Vazamento de dados
5. Não conseguir finalizar o protótipo
6. Melhor gestão de colaboladores
7. Diminuição de turnover
8. O grupo vai aprender como minerar e analizar dados
9. Baixo engajamento por parte dos envolvidos no projeto
10. Complexidade do projeto não mensurada
11. Atraso na disponibilização de recursos necessários a equipe do projeto
12. Defeitos no software podem não ser detectados até a sua implementação
13. Alta taxa de defeitos encontrados durante a homologação do produto pelo cliente
14. Qualidade do produto não atinge a expectativa do cliente
15. Equipe inexperiente para o nível de complexidade do projeto

4.1.6. Personas

Persona é uma representação real do cliente do produto que vai ser desenvolvido. A persona tem um nome, idade, hobbies, um trabalho e mostra para a empresa para quem eles devem desenvolver o produto e onde devem focar para que ele ajude os clientes da melhor maneira possível.

Meu grupo fez três personas, uma para o time de administração da empresa, outra para os funcionários da mesma e a última para o TechLead da empresa. Abaixo é possível visualizar nossas personas:

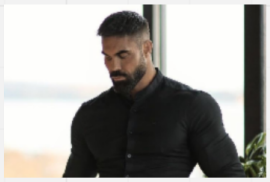
Neymar Júnior (Colaborador)

- 22 anos
- DEV júnior
- Insatisfeito com o salário
- Sem perspectiva de evolução da empresa
- Recebe muitas ofertas de trabalho
- Com o diagnóstico da predição, seria possível a empresa perceber a insatisfação e, com isso, entrar em contato com o colaborador
- Objetivo: Neymar é um morador de comunidade, a profissão de desenvolvedor abriu muitas portas para ele, com isso, Neymar deseja trabalhar em um ambiente saudável e ter uma boa remuneração para ajudar na renda de sua família.
- Dores: Neymar não ganha o tanto quanto ele gostaria, além de não ter uma perspectiva de evolução, fazendo com que sua rotina seja extremamente maçante, impossibilitando-o de aproveitar mais tempo com sua família. Devido a alta demanda de trabalho, ele desenvolveu algumas crises, como a ansiedade, sendo assim, ele também deseja um ambiente de trabalho mais receptivo.

Bianca Nepumoceno (time de RH)

- 28 anos
- Faz parte do time de RH
- Workaholic
- Extremamente preocupada com a saúde da empresa
- Está satisfeita com o seu trabalho
- Com o diagnóstico da predição, seria possível fazer o seu trabalho de forma mais eficiente
- Objetivo: Ela cresceu em uma família de classe média e nunca passou dificuldades, sendo assim, ela deseja se tornar a Head de Recursos Humanos, mas não tem pressa em alcançar esse objetivo.
- Dores: Perde muitos funcionários sem saber o motivo, e isso faz com que a empresa fique com uma reputação ruim, pois isso atrasa as entregas para os clientes, e mostra uma instabilidade. Ela deseja entender com mais propriedades quais são as insatisfações do colaborador.

Antony Vicente S. Maravalhas (Squad Leader)



- 36 Anos
- Jogador de golfe
- "Love is so short, forgetting is so long"
- É um dos pilares da empresa, consegue afirmar sua presença pelo seu carisma e comprometimento.
- Com o diagnóstico da predição, seria possível fazer o seu trabalho de forma mais eficiente.
- Objetivo: Quer deixar seu legado como líder, guiando sua equipe da melhor maneira possível
- Dores: Muitas vezes sente que seu time está desfalcado. Além disso, sempre entram pessoas novas em seu time devido à alta taxa de turnover, mas ele possui dificuldade para descobrir o motivo disso.

4.1.7. Jornadas do Usuário

A jornada de usuário consiste em um documento que apresenta, em ordem cronológica, o caminho que uma das personas criadas para o projeto precisa percorrer para concluir determinada tarefa relacionada ao problema.

Antony Vicente S. Maravalhas

Cenário: Antony está insastisfeito com desempenho um membro de sua equipe, pois ele não esta engajado e esta atrasando muitas entregas do projeto atual.

Expectativas

Conseguir engajar novamente o colaborador e reestruturar seu time

FASE 1 (Consultar o modelo)	FASE 2 (Entender o problema)	FASE 3 (Analisar o problema)	FASE 4 (Solucionar o problema)	FASE 5 (Retorno do rendimento)
1 - Consultar o algoritmo 2 - O funcionário em questão foi classificado como "propenso a sair"	1. Conversar com o colaborador para entender suas dores 2. O membro está agindo dessa forma pois não sentia que seu esforço estava sendo reconhecido.	1. Analisou a perfomance do funcionário na entrega dos ultimos projetos 2. Concluiu que o funcionário foi o diferencial em diversos projetos bem-sucedidos e, de fato, merecia reconhecimento.	1. Consultando no sistema, descobriu que o funcionário estava com uma promoção pendente há alguns meses. 2. Se convenceu de que o membro de sua equipe realmente merecia a promoção.	1. O funcionário recebe o reconhecimento que estava eperando 2. O engajamento da equipe volta ao normal e as entregas voltam a ser satisfatórias.

Oportunidades

Essa situação mostra como é importante reconheccer seus funcionários quando estão fazendo um bom trabalho para manter o foco da equipe.

Responsabilidades

Para aprimorar o reconhecimento dos colaboradores, a empresa poderia consultar o modelo preditivo com mais frequência, para ficar alerta de quais pessoas estão propensas a sair e poder tomar as devidas providências.

Neymar Júnior

Cenário: Neymar está insatisfeito com a falta de reconhecimento da empresa e está considerando sair já que recebe muitas ofertas de trabalho

Expectativas

Espera conseguir o reconhecimento qur deseja, pois já está acostumado com a vida e a cultura da empresa, e não deseja sair

FASE 1 (Comentar com um amigo suas impressões atuais)	FASE 2 (Squad Leader tenta entender o problema)	FASE 3 (Neymar tenta solucionar o problema)	FASE 4 (Análise do colaborador)	FASE 5 (Desfecho)
1 - Neymar comenta com um amigo de seu squad sobre sua instisfação. 2 - O amigo decide contar para o Squad Leader que seu colega está considerando sair da empresa, como uma forma de impedir que isso aconteça.	1. O Squad Leader conversa com Neymar para entender suas dores. 2. Neymar não se sente ouvido pelo Squad Leader.	1. Neymar vai até o RH para reclamar de seu chefe, pois ele não foi receptivo. 2. O time de RH disse que poderia conversar com o líder do squad sobre o ocorrido.	1. O Squad Leader, juntamente com o time de RH, analisaram o desempenho do Neymar e ainda confirmaram ele no modelo. 2. O modelo mostrava que ele estava "propenso a sair". 3. Com as análises, o Squad Leader concluiu que o Neymar não merecia o reconhecimento que havia pedido.	1. Neymar não ficou satisfeito com a resposta que recebeu do seu Squad Leader. 2. O funcionário então, aceita outra proposta de emprego.

Oportunidades

Considerando que o mercado está aquecido, muitos funcionários acabam super valorizando o seu trabalho, e as vezes exigem um maior reconhecimento do que apresentam nas entregas.

Responsabilidades

Essa situação mostra como o mercado de Dev's está aquecido e que se deve tomar cuidado nas contratações para conseguir reter a maior quantidade de funcionários possível.

4.2. Compreensão dos Dados

4.2.1 Descrição dos dados

Nesta seção, destrinchamos os dados fornecidos do arquivo de dados “Base Colaboradores Everymind_Inteli_2020 a 2022vModelo Preditivo” fornecida na plataforma Google Sheets pelo parceiro, a partir do documento no formato .XLSX. Por se tratar de informações confidenciais, todos os dados da Everymind não podem ser divulgados e devem ser restritos ao compartilhamento aberto.

1. **Everymind**

Nesta seção, destrinchamos os dados fornecidos da sessão “Everymind” fornecida. Suas formatações e significados são apresentados abaixo.

Essa seção é a base de dados dos colaboradores em um arquivo .XLSX, fonte Calibri, contendo informações de 475 funcionários (475 linhas) divididas em 14 colunas. Para cada colaborador foi disponibilizado: situação(ativo, desligado ou afastado), data de admissão, data de saída e tipo de saída(caso o funcionário não esteja mais na empresa), cargo, salário mês(último salário pago), data de nascimento, gênero, etnia, estado civil, escolaridade(grau), área (área de atuação), cidade, estado e idade.

Os dados estão mascarados, logo não aparece nome, cpf e nenhum outro dado que identifique o colaborador, com algumas exceções como o VP, pois só tem um na empresa. Os salários, também não são reais, mas seguem a proporcionalidade real para que possa ser usado na solução.

→ Variaveis

Elementos ↓

	Matrícula	Nome Completo	Data Admissao	Data Saida	Tipo Saida	Cargo	Salario Mês	Data Nascimento	Genero	Etnia	Estado Civil	Escolaridade	Estado	Cidade	Area
2	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
...	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
475	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
476	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

2. **Reconhecimento**

Nesta seção, destrinchamos os dados fornecidos da sessão “Reconhecimento” fornecida na plataforma Google Sheets pelo parceiro, a partir do documento no formato .XLSX. Suas formatações e significados são apresentados abaixo.

→ Variaveis

Elementos ↓

	Matricula	Codinome	Situação	Data de Admissão	Data Vigência	Novo Cargo	Novo Salario	Motivo	Alterou Função
2	X	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X	X
..	..	..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..	..	..
339	X	X	X	X	X	X	X	X	X
340	X	X	X	X	X	X	X	X	X

- Matrícula: A coluna contém o número de identificação do funcionário, definido por um número natural.
- Codinome: Esta coluna fornece o nome fictício do funcionário. O formato do dado é da forma “Pessoa Colaboradora ” seguida de um número natural.
- Situação: Descreve a situação atual do funcionário na empresa, podendo assumir estados: “Afastado”, “Ativo” ou “Desligado”.
- Data de Admissão: Informa a data de admissão do funcionário na empresa, no formato “MM/DD/AAAA”.
- Data Vigência: Define a data na qual o funcionário efetivamente começa a prestar seus serviços à empresa, apresentada no formato “DD/MM/AAAA”.
- Novo Cargo: Expressa o nome do novo cargo atribuído ao funcionário. Consiste em um texto que descreve a nova função. Na tabela fornecida, o valor da célula assume 26 valores. Exemplos: “Arquiteto Sr”, “Dev Jr”, "Líder IS”.
- Novo Salário: A coluna Novo Salário [sic] apresenta o valor do novo salário do funcionário. Assume o valor de um número com duas casas decimais.
- Motivo: Descreve a razão pela qual houve o remanejamento do cargo. O dado se apresenta na forma de um texto escrito em letras maiúsculas, tomando 3 (três) valores: “MÉRITO”, “PROMOÇÂO” e “RECLASSIF CARGO”.
- Alterou Função: Descreve se houve alteração de cargo, comparado ao anterior. O valor do dado é binário em texto, assumindo os valores “Sim” e “Não”.

3. Ambiente de Trabalho

Nesta seção, destrinchamos os dados fornecidos da sessão “Ambiente de Trabalho” fornecida na plataforma Google Sheets pelo parceiro, a partir do documento no formato .XLSX. Suas formatações e significados são apresentados abaixo.

A planilha de dados que foi disponibilizada é composta por colunas e linhas, cada linha contém os dados da pesquisa com um squad, ou seja, contendo um número de linhas iguais ao número de squad e as colunas contendo as variáveis estudadas.

Os dados dessa planilha são referentes ao estudo da Everymind de satisfação dos seus colaboradores no ambiente de trabalho, os dados são referentes a última pesquisa de satisfação realizada no dia 27/07/2022, e é feita entre todos os colaboradores de todos os setores a cada 3 meses.

➡ Variaveis

E

	Divisao	Pilar	Pontuação	Fator	Pontuação	Pergunta	Porcentagem das respostas	Taxa de Confiabilidade
2	X	X	X	X	X	X	X	X
3	X	X	X	X	X	X	X	X
...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...
1695	X	X	X	X	X	X	X	X

- Divisão - Informa o setor do squad que respondeu a pesquisa, o valor da variável é do tipo string e assume onze (11) valores, exemplo: Mkt Cloud, People & Culture e Vendas.
- Pilar - Categoria da pesquisa, o valor da variável é do tipo string e assume dez (10) valores, exemplo: Relacionamento com o gestor, Vestir a camisa e Crescimento pessoal.
- Pontuação - Pontuação referente ao pilar, o valor da variável é do tipo number e é definido por um número natural de 1 (um) a 10 (dez).
- Fator - Subcategoria do pilar, o valor da variável é do tipo string e assume vinte e sete (27) valores, exemplo: Confiança no gestor, Orgulho e Propósito e Direcionamento.
- Pontuação - Pontuação referente ao fator, o valor da variável é do tipo number e é definido por um número natural de 1 (um) a 10 (dez).
- Pergunta - Pergunta feita ao colaborador, o valor da variável é do tipo string e assume trinta e três (33) perguntas diferentes.
- Porcentagem das respostas - Informa a porcentagem das respostas por squad, a variável é do tipo string.
- Taxa de confiabilidade - Valor referente a credibilidade da resposta do squad, a variável é do tipo string.

4.2.1.1 Descrição do agrupamento e mescla

Todos os dados serão analisados individualmente e, a partir do grau de importância de suas inter relações, utilizamos relações matemáticas pertinentes a fim de estabelecer informações que auxiliem na construção da solução.

4.2.1.2 Descrição dos riscos e contingências

Em relação aos riscos e pertinências, podemos estabelecer a qualidade dos dados por critérios estabelecidos pelo grupo a partir do grau de importância para a solução, sua profundidade/superficialidade e a cobertura/diversidade (quantidade de informações que podem ser inferidas deles). O acesso aos dados é limitado aos que foram disponibilizados na planilha e, em eventuais situações, podem ser adicionados conforme o acordo feito entre o grupo e o parceiro.

4.2.1.3 Descrição dos criterios de escolha para análises iniciais

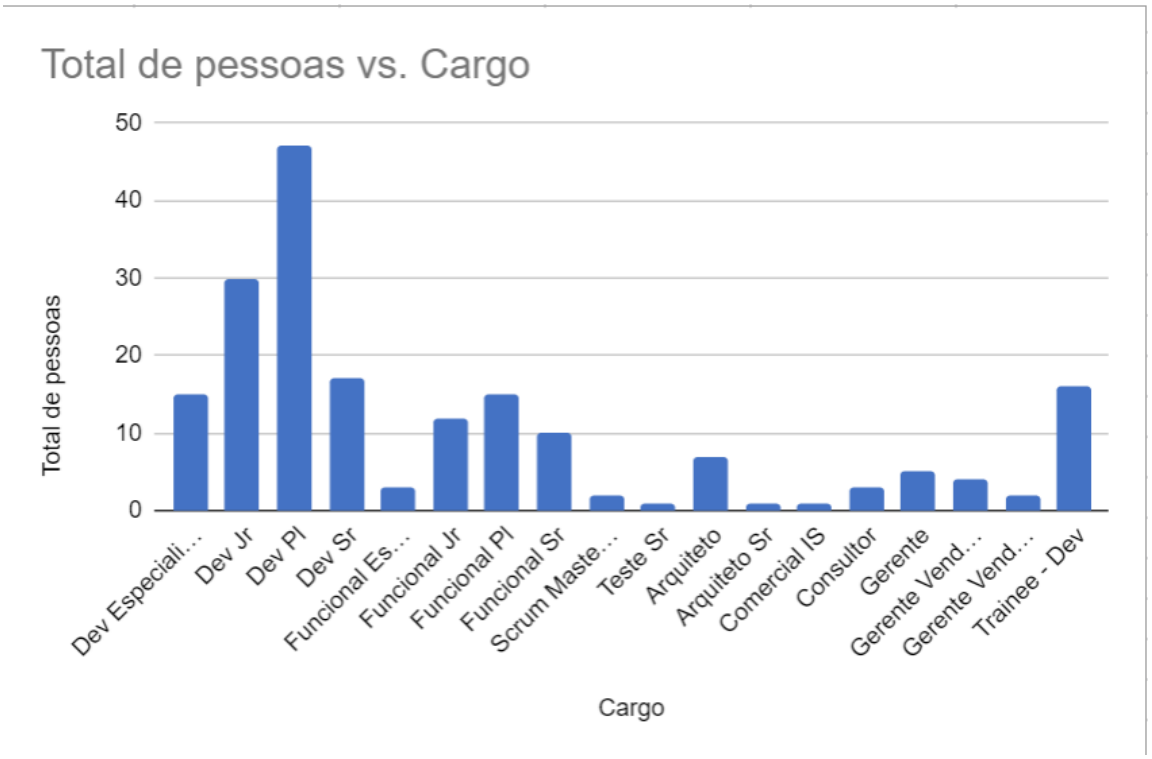
As análises iniciais foram baseadas a partir da relação cargo x saída da empresa, a partir disso criamos algumas hipóteses, utilizando esses dados para relacionar com os méritos/promoções e o salário.(*Seção 4.2.2*)

4.2.1.4 Descrição das restrições de segurança

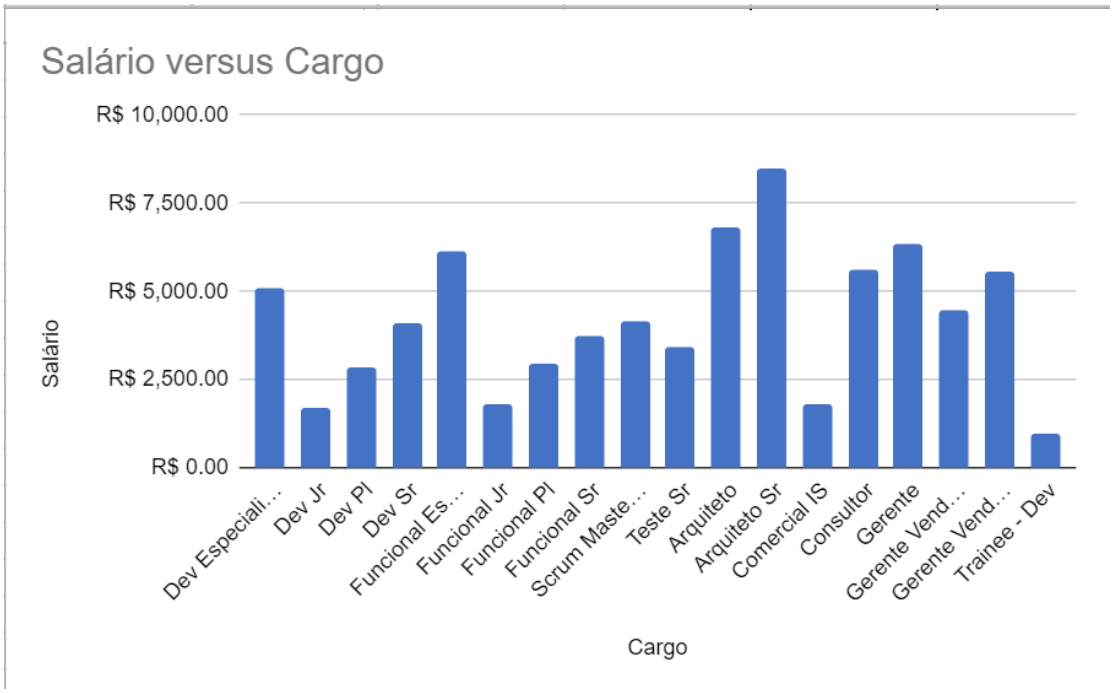
As bases de dados fornecidas não podem ser publicadas em nenhum lugar e os dados não devem ser divulgados.

4.2.2 Descrição estatística básica dos dados

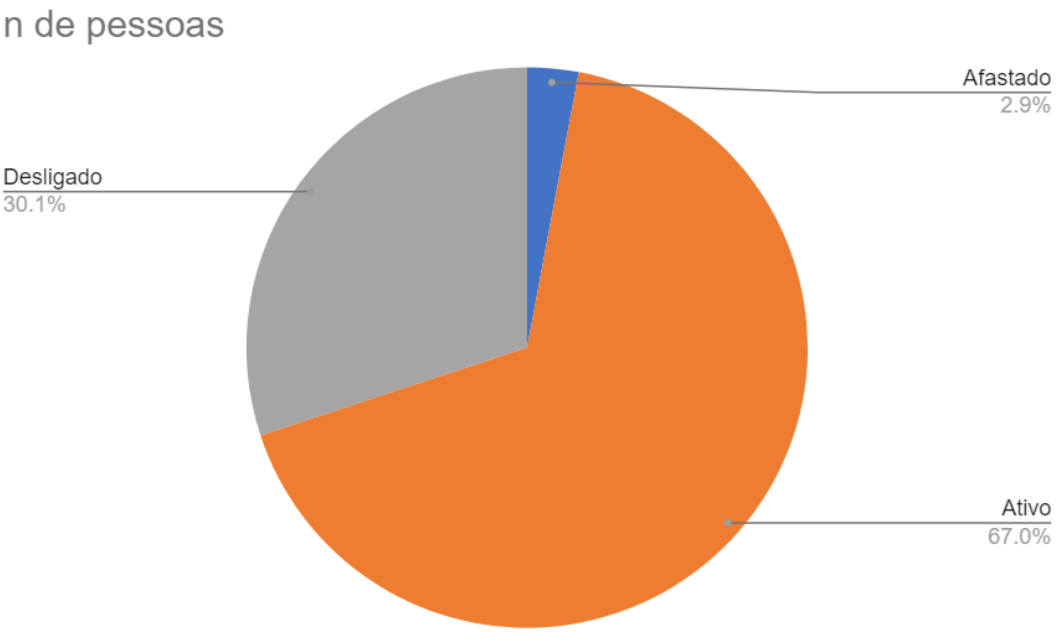
Descrição estatística básica dos dados, principalmente dos atributos de interesse, com inclusão de visualizações gráficas e como essas análises embasam suas hipóteses.



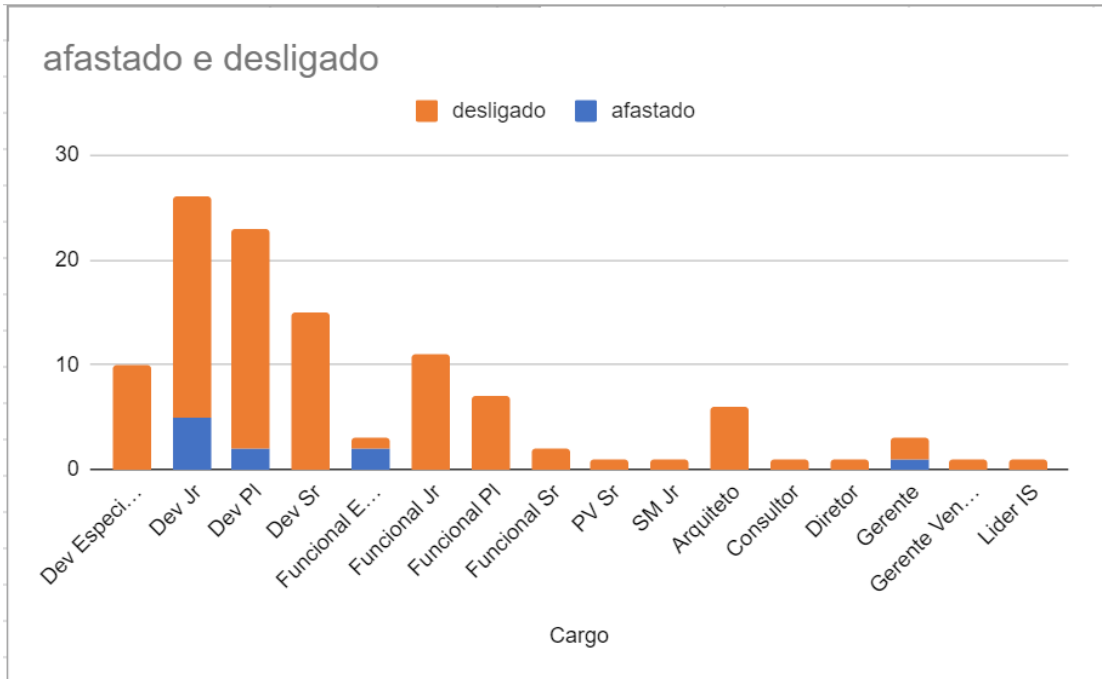
Esse gráfico mostra a relação do total de pessoas que foram demitidas ou pediram demissão e o cargo que elas exerciam.



Esse gráfico mostra a relação entre o salário das pessoas que foram demitidas ou pediram demissão e o cargo que elas exerciam.



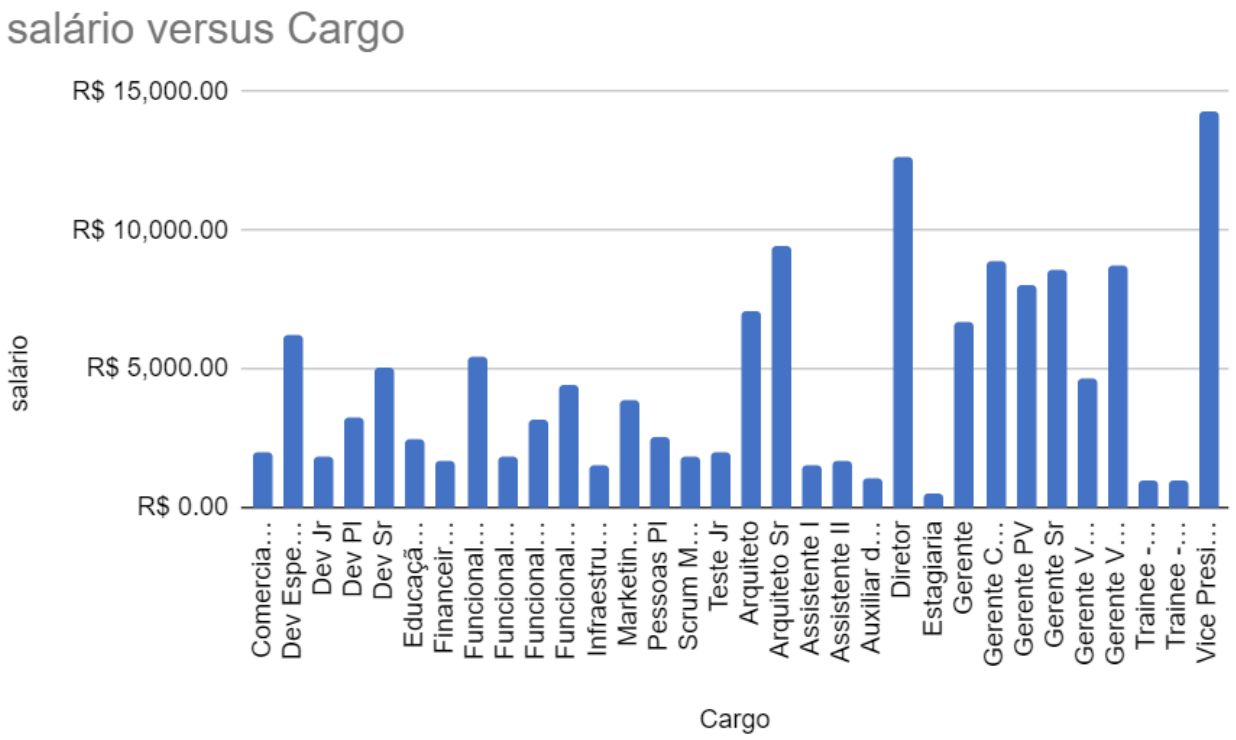
Esse gráfico mostra, das pessoas que receberam promoções (por mérito ou não), quais saíram da empresa e quais ainda trabalham lá.



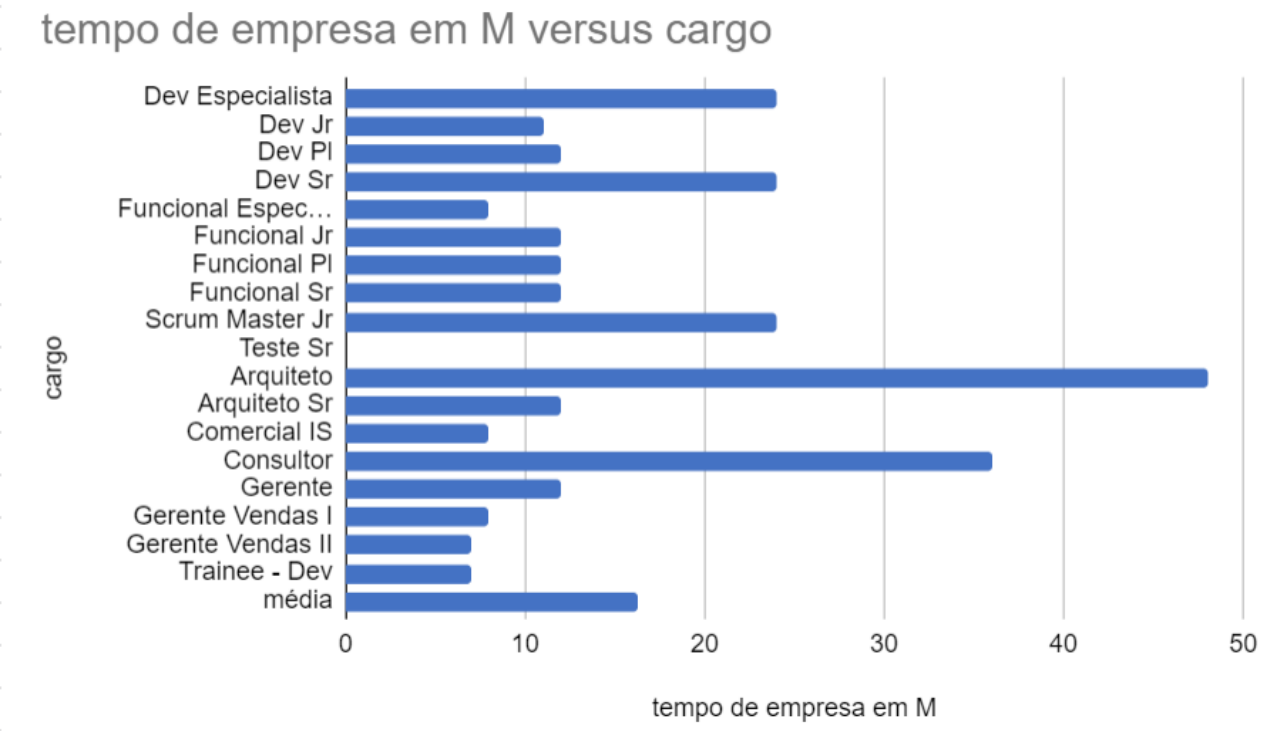
Esse gráfico mostra o cargo das pessoas que saíram da empresa, mas que receberam promoção (por mérito ou não).

Com a análise desses quatro gráficos, pode-se criar a hipótese de que os Devs da empresa não ficam nela por muito tempo, pois consideram o seu salário baixo, então no final eles sempre estão abertos a novas oportunidades que possam valorizar mais o seu trabalho.

Os cargos de Devs são os que mais possuem pessoas trabalhando, mas também os que mais tem demissões, o que por hora faz sentido, mas analisando os dados pode-se perceber que as proporções são interessantes. Por exemplo com



Esse gráfico mostra a relação do salário das pessoas que ainda trabalham na empresa com o cargo que exercem.



Esse gráfico mostra o tempo médio de cada cargo em meses.

Com os dois gráficos acima podemos inferir que a retenção dos funcionários na empresa é baixa. Nossa hipótese é que isso acontece, pois os salários podem estar abaixo do que o mercado oferece.

4.2.3 Descrição da predição desejada

O modelo de predição será classificatório, ou seja, as classes já estão pré-definidas e o algoritmo vai definir qual colaborador se encaixa em cada classe. A natureza do modelo(binária ou múltiplas classes) ainda não foi definida.

4.3. Preparação dos Dados

A manipulação dos dados exige que eles estejam todos em formato de número (type: number) para fácil leitura e carregamento dos dados pelo algoritmo, os dados disponibilizados precisam ser passados por uma etapa de preparação. Essa etapa inclui tarefas de classificação e formatação de dados para modelagem, remover ou substituir registros em branco, seleção de um subconjunto de amostras para análise, derivação de novos atributos e mesclar conjuntos de dados e registros.

4.3.1 Classificação e formatação de dados para modelagem

Para a formatação inicialmente precisaremos tirar os espaços de toda a tabela a fim de padronizar todos os dados de todas as colunas, para isso fizemos a substituição de espaço (“ ”) para (“”), exemplo: “Superior incompleto” para “Superiorincompleto”. Essa Feature foi selecionada para possibilitar a utilização do Label Encoder e do One Hot Encoder para a categorização das informações. Um exemplo das transformações segue abaixo:

0	Dev Pl
1	Dev Pl
2	Dev Pl
3	Dev Sr
4	Funcional Sr
5	Teste Sr
6	Dev Pl
7	Dev Especialista
8	Dev Pl
9	Dev Pl
10	Dev Especialista

0	DevPl
1	DevPl
2	DevPl
3	DevSr
4	FuncionalSr
5	TesteSr
6	DevPl
7	DevEspecialista
8	DevPl
9	DevPl
10	DevEspecialista

Durante esse processo vamos tratar os dados a fim de padronizá-los para que sejam aceitos e melhor utilizados pelo algoritmo a partir de funções que modificam a forma do dado. No momento, estamos trabalhando com alguns tipos de dados, sendo eles dados relacionados a tempo e dados relacionados a nome.

Nos dados relacionados à data foram formatados apenas a ordem de dd/mm/yyyy (Exemplo: 31/10/2003) para yyyy/mm/dd (Exemplo: 2003/10/31). Sofrem essa alteração os dados presentes na aba “Everymind” nas colunas "Dt Admissao", "Dt Nascimento" e "Dt Saida" e na aba “Reconhecimento” nas colunas “Data de Admissão” e “Data Vigência”. Essa Feature foi selecionada para padronizar todas as datas do banco e facilitar o cálculo entre duas datas para análises futuras. Um exemplo das transformações segue abaixo:

0	09-04-1985
1	24-03-1998
2	28-01-1985
3	24-03-1995
4	24-01-1989
5	26-01-1990
6	16-09-2000
7	19-10-1981
8	29-06-1986
9	14-08-1991
10	27-11-1995

0	1985-04-09
1	1998-03-24
2	1985-01-28
3	1995-03-24
4	1989-01-24
5	1990-01-26
6	2000-09-16
7	1981-10-19
8	1986-06-29
9	1991-08-14
10	1995-11-27

Nos dados relacionados a nome foram formatados os textos com o objetivo de permanecer apenas os números. Exemplo: “PessoaColaboradora197” foi formatado para apenas “197”. Sofrem essa alteração os dados presentes na aba “Everymind” nas coluna “Nome Completo” e na aba “Reconhecimento” na coluna “Codinome”. Essa Feature foi selecionada para padronizar todos os dados categóricos em números e facilitar análises futuras do algoritmo.
Um exemplo das transformações segue abaixo:

0	Pessoa Colaboradora 1
1	Pessoa Colaboradora 10
2	Pessoa Colaboradora 100
3	Pessoa Colaboradora 101
4	Pessoa Colaboradora 102
5	Pessoa Colaboradora 103
6	Pessoa Colaboradora 104
7	Pessoa Colaboradora 105
8	Pessoa Colaboradora 106
9	Pessoa Colaboradora 107
10	Pessoa Colaboradora 108

0	1
1	10
2	100
3	101
4	102
5	103
6	104
7	105
8	106
9	107
10	108

As colunas “Pulou”, “Muito Insatisfeito”, “Insatisfeito”, “Neutro”, “Satisfeito” e “Muito Satisfeito” da aba “Ambiente de Trabalho 27.07” mesmo estando em porcentagem o algoritmo reconhece como formato de texto (String) e para transformar em número (Number) trocando os (“%”) por (“”). Exemplo: “45,67%” foi transformado apenas para “45,67”. Essa Feature foi selecionada para padronizar todos os dados do banco no tipo número e facilitando o manuseio para análises futuras. Um exemplo das transformações segue abaixo:

0	77.97%
1	67.06%
2	62.5%
3	48.96%
4	50%
5	50%
6	NaN
7	NaN
8	66.67%
9	81.45%
10	80%

0	77.97
1	67.06
2	62.5
3	48.96
4	50
5	50
6	NaN
7	NaN
8	66.67
9	81.45
10	80

Grande parte dos dados são categóricos e quando temos categorias como descrição do dado precisamos converter para valores numéricos, podemos fazer isso usando o Label Encoder que faz uma atribuição numérica crescente para cada categoria impondo uma ordenação entre as classes e o One Hot Encoder que cria uma coluna para cada valor e faz uma atribuição do valor 1(um) para a coluna correspondente da amostra e consequentemente não necessitando de uma ordenação.
As transformações usando o Label Encoder foi usado na tabela “Everymind” na coluna “Escolaridade ”, fizemos uma atribuição numérica dos dados da coluna em ordem crescente e ordinal. Exemplo: “EnsinoMédioIncompleto” atribui “0”, “EnsinoMédio” atribui “1”, etc. Essa Feature foi selecionada para transformar os dados categóricos em numéricos possibilitando a utilização deles em análises futuras . Um exemplo das transformações segue abaixo:

0	Superior incompleto
1	Graduação
2	Pós Graduação
3	Graduação
4	Graduação
5	Graduação
6	Graduação
7	Graduação
8	Graduação
9	Graduação
10	Graduação

0	3
1	4
2	5
3	4
4	4
5	4
6	4
7	4
8	4
9	4
10	4

As transformações usando o One Hot Encoder foram usadas em todas as tabelas do banco de dados e em dezessete colunas no total, foi feito uma atribuição dos valores em colunas e uma atribuição de números (0 e 1) à essa colunas para indicar se a coluna é correspondente a amostra, exemplo: Os valores da coluna “Estado Civil” se transformaram em colunas e foi atribuído o número 1 (um) para correspondente e 0 (zero) para não correspondente, podendo só ter apenas um número 1(um) na linha. Essa Feature foi selecionada para transformar os dados categóricos em numéricos possibilitando a utilização deles em análises futuras. Um exemplo das transformações segue abaixo:

0 Casado	0 0
1 Solteiro	1 1
2 Solteiro	2 1
3 Solteiro	3 1
4 Casado	4 0
5 Solteiro	5 1
6 Solteiro	6 1
7 Casado	7 0
8 Casado	8 0
9 Solteiro	9 1
10 Solteiro	10 1

4.3.2 Remover ou substituir registros em branco

Em nosso modelo preditivo ter registros em brancos prejudica a análise do algoritmo, tendo isso em vista, detectamos que nas colunas “Dt Saida” e “Tipo Saida” da aba “Everymind” e as colunas “Pulou”, “Muito Insatisfeito", “Insatisfeito”, “Neutro”, “Satisfeito” e “Muito Satisfeito” da aba “Ambiente de Trabalho 27.07” haviam dados vazios e precisariam ser preenchidos. Essa Feature foi selecionada para os campos vazios na tabela não ocasionarem erros em nossa predição do algoritmo e prejudicar a confiabilidade das informações.

Da aba “Everymind”, a coluna “Dt Saida” estavam em formato de data e os valores vazios representavam que o colaborador daquela linha em específico ainda estava ativo na empresa, então apenas substituímos o valor de nulo para a data atual que se atualiza conforme os dias passam. A coluna “Tipo Saida” estava em formato de texto(string) e os valores vazios na coluna representam que o colaborador ainda está ativo na empresa, então substituímos o valor nulo para “ColaboradorAtivo”. Um exemplo das transformações segue abaixo:

464	NaN
465	NaN
466	NaN
467	NaN
468	NaN
469	NaN
470	NaN
471	NaN
472	NaN
473	NaN
474	NaN

464	ColaboradorAtivo
465	ColaboradorAtivo
466	ColaboradorAtivo
467	ColaboradorAtivo
468	ColaboradorAtivo
469	ColaboradorAtivo
470	ColaboradorAtivo
471	ColaboradorAtivo
472	ColaboradorAtivo
473	ColaboradorAtivo
474	ColaboradorAtivo

Da aba “Ambiente de Trabalho 27.07”, como se tratavam de dados numéricos e os valores vazios representavam que aquela opção não foi escolhida por nenhum colaborador do setor que participou da pesquisa, então apenas substituímos o valor de vazio para o número 0 (zero). Um exemplo das transformações segue abaixo:

0	NaN
1	NaN
2	NaN
3	1.04%
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	0.36%
10	NaN

0	0
1	0
2	0
3	1.04%
4	0
5	0
6	0
7	0
8	0
9	0.36%
10	0

4.3.3 Seleção de um subconjunto de amostras para análise

Dentro da tabela, fizemos uma seleção de amostra de todos os funcionários que saíram da empresa e criamos uma nova tabela apenas com esses dados. Essa Feature foi selecionada para dar um foco nos colaboradores inativos, facilitando a análise e decisão de quais fatores mais influenciam a decisão de deixar a empresa.
Funcionários que saíram em menos de um ano. Essa Feature foi selecionada para dar um foco nos colaboradores que saíram da empresa em menos de um ano dentre os inativos para investigar as variáveis que mais influenciam na decisão.

4.3.4 Derivação de novos atributos

Durante toda a formatação dos dados, foi detectada a necessidade de cálculos entre as datas da tabela. Essa Feature foi selecionada para facilitar a análise dos dados, deixando de lado a necessidade de fazer cálculos complexos com frequência.
Na aba “Everymind” fizemos o cálculo entre a data de admissão (“Dt Admissao”) e a data de saída (“Dt Saida”) para obter os meses de empresa e entre a data de nascimento (“Dt Nascimento”) e a data de saída (“Dt Saida”) para obter a idade dos colaboradores. Um exemplo dos cálculos segue abaixo:

0	2020-04-13	2020-05-12
1	2020-07-01	2020-08-07
2	2020-07-01	2020-09-21
3	2020-08-03	2020-10-30
4	2020-08-10	2020-09-18
5	2020-09-01	2020-09-17

0	29
1	37
2	82
3	88
4	39
5	16
6	83
7	74
8	45
9	4
10	79

Na aba “Reconhecimento” fizemos o cálculo entre a data de admissão (“Data de Admissão”) e a data de vigência (“Data Vigência”) para obter o número de dias entre a data de admissão e a data que o colaborador foi reconhecido na empresa. Um exemplo dos cálculos segue abaixo:

0	2022-02-14	2022-07-01
1	2019-12-02	2022-02-01
2	2019-12-02	2021-08-01
3	2019-12-02	2021-06-01
4	2021-11-03	2022-06-01
5	2021-11-03	2022-03-01

```
0     137
1     792
2     608
3     547
4     210
5     118
6     879
7     818
8     361
9     323
10    295
```

4.3.5 Colunas não utilizadas

Nesta etapa foi feita uma seleção dos dados e definição da relevância do atributo no nosso modelo, um atributo observado foi a questão da etnia, usar isso no modelo como fator decisivo é antiético, cria um viés negativo, deixa o modelo questionável com nível baixo de credibilidade e tira a viabilidade da solução, pensado nisso foi retirado da análise a coluna contendo a etnia dos colaboradores. Essa feature foi selecionada para analisar as variáveis que não tem validade para o resultado e restringi-la especificamente, já que não fazem sentido para o negócio.

4.3.6 Mesclar conjuntos de dados e registros

Ao fim da formatação, categorização e padronização do banco de dados, todas as informações foram transferidos para uma nova tabela em que o algoritmo poderá trabalhar com ela no backend. Foi criada uma nova tabela correspondente para cada aba da base de dados. Essa Feature foi selecionada para que o algoritmo possa fazer as análises em uma tabela com os dados formatados sem que isso altere a tabela original da empresa.

4.4. Modelagem

Para a Sprint 3, você deve descrever aqui os experimentos realizados com os modelos (treinamentos e testes) até o momento. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

Para a Sprint 4, você deve realizar a descrição final dos experimentos realizados (treinamentos e testes), comparando modelos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus experimentos e resultados.

4.5. Avaliação

Nesta seção, descreva a solução final de modelo preditivo, e justifique a escolha. Alinhe sua justificativa com a seção 4.1, resgatando o entendimento do negócio e explicando de que formas seu modelo atende os requisitos. Não deixe de usar equações, tabelas e gráficos de visualização de dados para melhor ilustrar seus argumentos.

4.6 Comparação de Modelos

5. Conclusões e Recomendações

Escreva, de forma resumida, sobre os principais resultados do seu projeto e faça recomendações formais ao seu parceiro de negócios em relação ao uso desse modelo. Você pode aproveitar este espaço para comentar sobre possíveis materiais extras, como um manual de usuário mais detalhado na seção “Anexos”.

Não se esqueça também das pessoas que serão potencialmente afetadas pelas decisões do modelo preditivo, e elabore recomendações que ajudem seu parceiro a tratá-las de maneira estratégica e ética.

6. Referências

Nesta seção você deve incluir as principais referências de seu projeto, para que seu parceiro possa consultar caso ele se interessar em aprofundar.

Utilize a norma ABNT NBR 6023 para regras específicas de referências. Um exemplo de referência de livro:

SOBRENOME, Nome. **Título do livro**: subtítulo do livro. Edição. Cidade de publicação: Nome da editora, Ano de publicação.

Anexos

Utilize esta seção para anexar materiais como manuais de usuário, documentos complementares que ficaram grandes e não couberam no corpo do texto etc.