



Faculdade de Design,
Tecnologia e Comunicação
 Universidade Europeia

O fenómeno da rotatividade sob a ótica dos Recursos Humanos

Trabalho realizado no âmbito da cadeira de
Projeto e Visualização de Informação do docente
Alessandro Pinheiro, em conjunto com as
cadeiras de Business Intelligence; Data Mining e
Web Analytics, Prof. Márcio Saraiva; Economia e
Sistemas de Informação, Prof. João R. Santos;

Trabalho realizado por Ana Cristina Jesus

Informática de Gestão

2023/24

Resumo: O relatório aborda a análise do atrito de funcionários em uma empresa, utilizando dados fictícios provenientes do Kaggle e elaborados pela IBM. A metodologia incluiu tratamento de dados, pré-processamento, análise descritiva e preditiva, com ênfase em regras de associação e modelos Random Forest. Além disso, foi incorporada uma análise geográfica por meio de Sistemas de Informação Geográfica (SIG). O estudo concluiu que fatores demográficos, de desempenho, satisfação e experiência influenciam o atrito, sendo a proximidade geográfica menos impactante. Implicações econômicas foram discutidas, relacionando a alta rotatividade com produtividade, lucro, custos de recrutamento e imagem corporativa, com uma perspectiva utilitarista.

Palavras-Chave: Recursos Humanos, RH, análise, organização, colaboradores, Atrito, dados, funcionários, empresa, rotatividade, utilitarista, produtividade, custos de recrutamento, Imagem Corporativa

Introdução

A Análise de Recursos Humanos (Análise de RH) está na vanguarda da gestão organizacional moderna, representando uma abordagem transformadora para compreender e otimizar a força de trabalho. À medida que navegamos numa era inundada de dados, a Análise de RH surge como uma ferramenta poderosa que capacita as organizações a desvendar as complexidades do seu capital humano, especialmente no domínio do desgaste – onde a rotatividade de funcionários representa um desafio crítico.

A base utilizada para este estudo, fictícia e criada pela IBM, destaca-se como um elemento crucial na abordagem da Análise de RH. Frequentemente utilizada em estudos acadêmicos, essa base visa identificar os principais fatores que influenciam a decisão de um funcionário de deixar uma organização.

Basicamente, a Análise de RH envolve a análise sistemática de vastos conjuntos de dados organizacionais para extrair insights significativos relacionados aos recursos humanos. Ele investiga várias facetas, desde dados demográficos e indicadores de desempenho dos funcionários até níveis de satisfação no trabalho e preditores de saídas voluntárias. O objetivo geral é equipar os profissionais de RH e os tomadores de decisão com insights baseados em dados que lhes permitam tomar decisões informadas, implementar estratégias direcionadas e alinhar as práticas de recursos humanos com objetivos de negócios mais amplos.

O estudo da análise de RH, especialmente no contexto de desgaste, tem um significado profundo por várias razões convincentes. O principal é o imperativo de enfrentar os desafios multifacetados colocados pela rotatividade de funcionários. O atrito, com os custos e interrupções associados, exige uma abordagem proativa e estratégica. A Análise de RH fornece as ferramentas necessárias para dissecar as taxas de desgaste, identificar os fatores subjacentes e formular intervenções que promovam uma força de trabalho estável e engajada.

Num cenário onde o talento é um diferencial crítico, as organizações reconhecem que manter uma força de trabalho coesa e qualificada é fundamental para o sucesso sustentado. É aqui que reside a relevância da Análise de RH: ela oferece uma lente através da qual as organizações podem não apenas compreender tendências históricas, mas também prever e antecipar futuros casos de desgaste. Esta capacidade preditiva é particularmente crucial para o pessoal de RH que

procura reter talentos valiosos e minimizar os impactos financeiros e operacionais da rotatividade.

O estudo da Análise de RH transcende a mera análise de dados; incorpora uma mudança de paradigma na gestão de recursos humanos. Ao aproveitar a tecnologia, os modelos estatísticos e os algoritmos de aprendizado de máquina, as organizações obtêm uma compreensão mais profunda da intrincada rede de fatores que influenciam o desgaste. Vai além de medidas reativas, capacitando os profissionais de RH a adotarem uma postura proativa, antecipando desafios e implementando estratégias de retenção direcionadas.

Além disso, a Análise de RH alinha-se com o espírito mais amplo da tomada de decisões baseada em evidências. Numa era em que as organizações são inundadas com dados, a capacidade de destilar insights acionáveis a partir destas informações torna-se uma vantagem competitiva. O estudo de Análise de RH capacita estudantes universitários de gestão, especialmente aqueles no último ano, com um conjunto de habilidades que não só é muito procurado, mas também os posiciona como contribuidores estratégicos para o sucesso organizacional.

À medida que as organizações enfrentam as complexidades de um ambiente de negócios dinâmico, compreender e aplicar a Análise de RH torna-se a base de uma gestão eficaz. É uma abordagem holística que considera não apenas os aspectos quantitativos do desgaste, mas também as dimensões qualitativas, como a satisfação e o bem-estar dos funcionários. Em essência, o estudo de Análise de RH convida os estudantes de gestão a abraçar um futuro onde as decisões não são apenas guiadas pela intuição, mas estão enraizadas nos insights empíricos derivados da análise de dados de capital humano.

Nesta exploração abrangente de Análise de RH e sua relevância para o atrito, os estudantes de administração embarcam em uma jornada que transcende as fronteiras tradicionais. É uma jornada marcada pela inovação, pela visão estratégica e pelo compromisso de nutrir e reter o ativo mais valioso de qualquer organização: o seu pessoal.

Sobre a Base de Dados, e o processo de tratamento utilizado

No âmbito deste projeto, a base de dados foi extraída do Kaggle, embora a sua origem veio da empresa IBM. A empresa reuniu informações sobre satisfação dos funcionários, renda, antiguidade e alguns dados demográficos. Inclui os dados de 1.470 funcionários.

O primeiro passo do processo ETL consistiu na identificação de colunas dispensáveis, tais como "Over 18," uma vez que a idade mínima já era de 18 anos, assim como as colunas "Standard Hours" e "Employee Count," que apresentavam apenas valores únicos e, portanto, não contribuíam para a análise.

Posteriormente, realizou-se uma análise cuidadosa em busca de valores nulos, em falta ou repetidos. Contudo, concluiu-se que a base de dados estava isenta dessas irregularidades. Para otimizar a análise, efetuou-se uma síntese dos valores únicos em cada coluna.

Além disso, algumas colunas foram transformadas para facilitar a compreensão e interpretação dos dados. As colunas que originalmente continham valores em formato numérico foram convertidas em categorias, o que aprimorou a visualização e entendimento das informações.

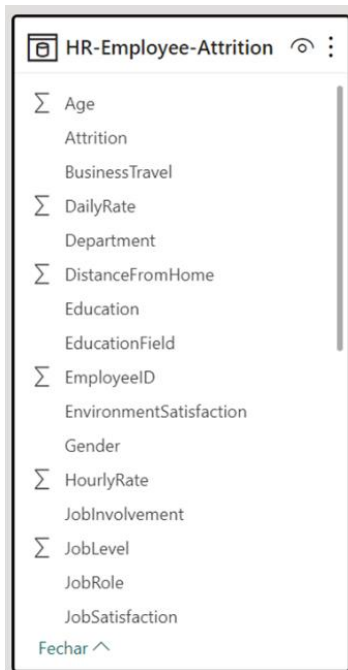
Segue abaixo uma descrição aprimorada das colunas, destacando a transformação realizada em algumas delas:

Name	Description	MONTHLY INCOME	Numerical Value - MONTHLY SALARY
AGE	Numerical Value	MONTHLY RATE	Numerical Value - MONTHLY RATE
ATTRITION	Employee leaving the company (0=no, 1=yes)	NUMCOMPANIES WORKED	Numerical Value - NO. OF COMPANIES WORKED AT
BUSINESS TRAVEL	(1=No Travel, 2=Travel Frequently, 3=Travel Rarely)	OVER 18	(1=YES, 2=NO)
DAILY RATE	Numerical Value - Salary Level	OVERTIME	(1=NO, 2=YES)
DEPARTMENT	(1=HR, 2=R&D, 3=Sales)	PERCENT SALARY HIKE	Numerical Value - PERCENTAGE INCREASE IN SALARY
DISTANCE FROM HOME	Numerical Value - THE DISTANCE FROM WORK TO HOME	PERFORMANCE RATING	Numerical Value - ERFORMANCE RATING
EDUCATION	Numerical Value	RELATIONS SATISFACTION	Numerical Value - RELATIONS SATISFACTION
EDUCATION FIELD	(1=HR, 2=LIFE SCIENCES, 3=MARKETING, 4=MEDICAL SCIENCES, 5=OTHERS, 6=TEHCNICAL)	STANDARD HOURS	Numerical Value - STANDARD HOURS
EMPLOYEE COUNT	Numerical Value	STOCK OPTIONS LEVEL	Numerical Value - STOCK OPTIONS
EMPLOYEE NUMBER	Numerical Value - EMPLOYEE ID	TOTAL WORKING YEARS	Numerical Value - TOTAL YEARS WORKED
ENVIROMENT SATISFACTION	Numerical Value - SATISFACTION WITH THE ENVIROMENT	TRAINING TIMES LAST YEAR	Numerical Value - HOURS SPENT TRAINING
GENDER	(1=FEMALE, 2=MALE)	WORK LIFE BALANCE	Numerical Value - TIME SPENT BEWTWEEN WORK AND OUTSIDE
HOURLY RATE	Numerical Value - HOURLY SALARY	YEARS AT COMPANY	Numerical Value - TOTAL NUMBER OF YEARS AT THE COMPNAY
JOB INVOLVEMENT	Numerical Value - JOB INVOLVEMENT	YEARS IN CURRENT ROLE	Numerical Value -YEARS IN CURRENT ROLE
JOB LEVEL	Numerical Value - LEVEL OF JOB	YEARS SINCE LAST PROMOTION	Numerical Value - LAST PROMOTION
JOB ROLE	(1=HC REP, 2=HR, 3=LAB TECHNICIAN, 4=MANAGER, 5= MANAGING DIRECTOR, 6= REASEARCH DIRECTOR, 7= RESEARCH SCIENTIST, 8=SALES EXECUTIEVE, 9= SALES REPRESENTATIVE)	YEARS WITH CURRENT MANAGER	Numerical Value - YEARS SPENT WITH CURRENT MANAGER
JOB SATISFACTION	Numerical Value - SATISFACTION WITH THE JOB		

Foi feita então o estudo a valores em falta, nulos, ou repetidos, onde se concluiu que não era o caso nesta tabela. E assim, procedemos à sua análise e criação de visualizações.

Business Intelligence

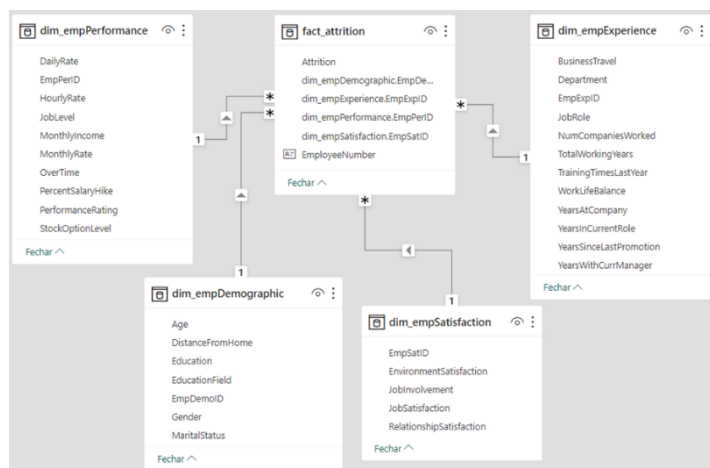
Ao extrair a base de dados para o Power BI, inicialmente, não foram identificados valores nulos ou repetidos. Contudo, uma análise mais aprofundada revelou a presença de colunas menos relevantes, como "Standard Hours", "Over 18" e "Employee Count". Essas colunas, devido à sua escassa variedade de valores, demonstraram ter pouca contribuição informativa para a análise em questão.



Nesse contexto, procedeu-se à revisão da estrutura da base de dados, visando otimizar a relevância das informações. As colunas mencionadas foram removidas para simplificar o conjunto de dados, focando em atributos mais pertinentes para a análise da taxa de atrito. Dentre os atributos retidos, destacam-se informações cruciais, como idade, género, grau de educação, satisfação no trabalho, experiência profissional, tempo na empresa atual, entre outros.

Com o intuito de aprimorar a organização e a utilidade na visualização e análise dos dados, optou-se por distribuir esses atributos em quatro tabelas distintas: Demografia, Satisfação, Experiência e Desempenho. Essa abordagem facilita a criação de visualizações mais específicas e a execução de queries direcionadas, proporcionando uma compreensão mais profunda dos fatores que influenciam a taxa de atrito.

A tabela "Demografia" concentra informações relacionadas à composição demográfica dos colaboradores, como idade, género e grau de educação. A tabela "Satisfação" engloba dados sobre a satisfação no trabalho, um fator crucial na retenção de talentos. A tabela "Experiência" abrange informações como tempo na empresa atual e experiência profissional, enquanto a tabela "Desempenho" destaca aspetos ligados ao desempenho dos colaboradores.



Essa reorganização não apenas simplifica a manipulação dos dados no Power BI, mas também oferece uma estrutura mais intuitiva para explorar e compreender as dinâmicas subjacentes à taxa de atrito. A segmentação das informações em tabelas específicas permite uma análise mais granular e facilita a identificação de padrões e tendências que podem direcionar estratégias de retenção de recursos humanos.

I- Criação de Medidas DAX

A linguagem DAX desempenha um papel fundamental no ambiente do Power BI, semelhante à sua utilização no Excel. Essa linguagem permite a criação de tabelas personalizadas, medidas e colunas calculadas, constituindo-se em um recurso robusto para a análise de dados. No âmbito deste contexto específico, a aplicação da DAX foi direcionada para calcular a taxa de atrito, uma métrica crucial para compreender a dinâmica de saída de colaboradores na empresa.

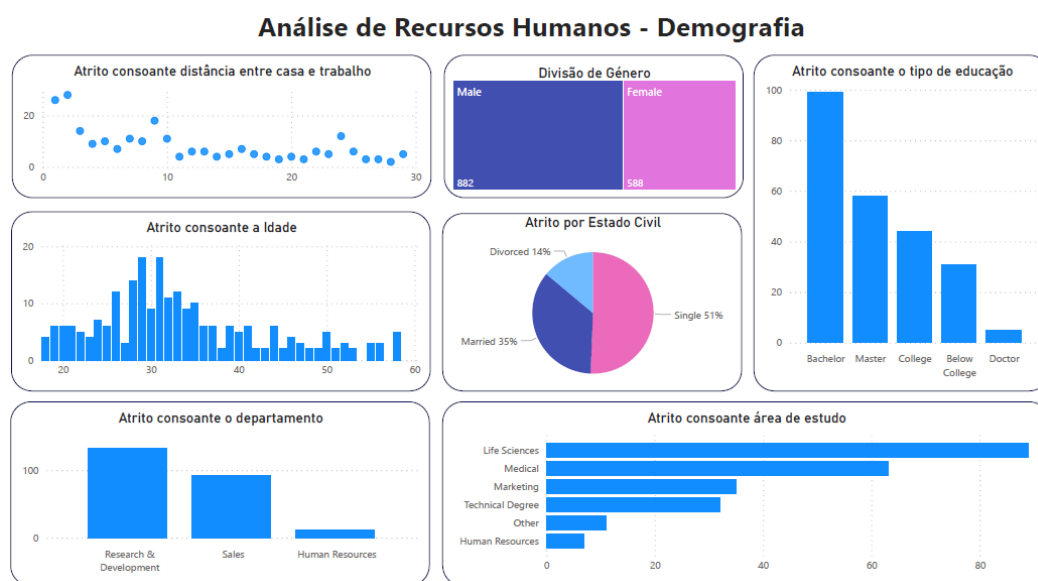
```
1 AttritionCount = CALCULATE(COUNTROWS(fact_attrition), fact_attrition[Attrition] = "Yes")
1 AttritionRate = [AttritionCount]/COUNT(fact_attrition[EmpID])
1 MédiaSalárioAnual = 12 * DIVIDE( SUM(dim_empPerf[MonthlyIncome]), DISTINCTCOUNT(dim_empPerf[EmpPerID]))
```

Dentro desse contexto, algumas medidas essenciais foram desenvolvidas para aprofundar o estudo do atrito. A medida "Attrition Rate" apresenta uma fórmula de turnover adaptada aos dados disponíveis na base de dados, proporcionando uma visão perspicaz da taxa de saída de colaboradores. A medida "Attrition Count" contabiliza a quantidade de indivíduos que deixaram a empresa, fornecendo uma perspectiva quantitativa dessa movimentação de pessoal. Além disso, a medida "Média Salário Anual" oferece uma estimativa do salário anual dos colaboradores, contribuindo para a análise contextualizada do atrito.

Essas medidas não apenas desempenham um papel fundamental na avaliação do atrito, mas também servem como alicerce para a criação de visualizações significativas. Ao compreender e utilizar essas métricas de maneira integrada, é possível obter insights valiosos que facilitam a interpretação dos dados de forma abrangente. Essa abordagem estruturada não apenas enriquece a compreensão do atrito, mas também fornece uma base sólida para análises mais aprofundadas e tomadas de decisão informadas

II- Criação de Dashboards

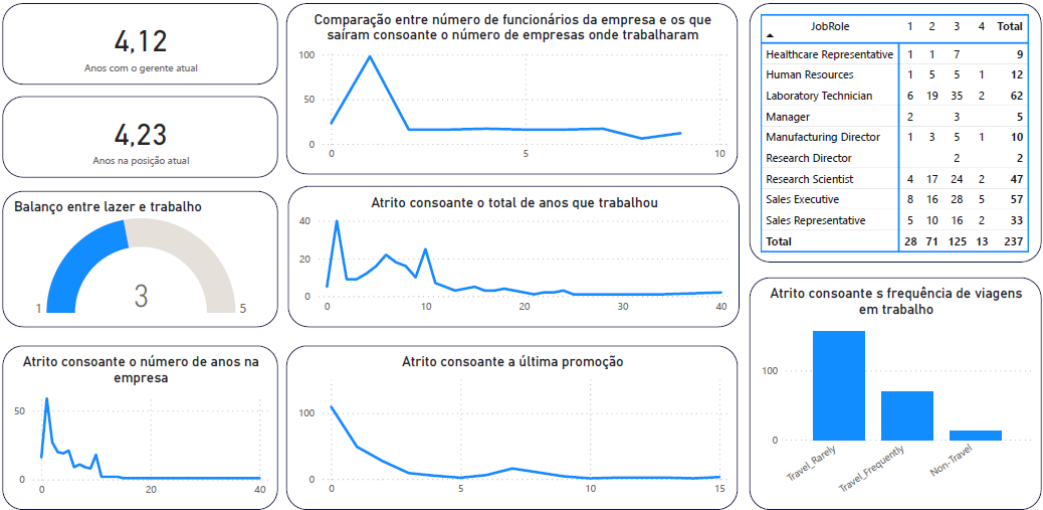
1. Demografia do Atrito



As visualizações foram elaboradas com base nos dados relativos aos colaboradores que deixaram a empresa, proporcionando a oportunidade de analisar os principais fatores que influenciaram a maioria das saídas. A partir dessas análises, foi possível identificar um padrão de maior atrito entre colaboradores do sexo masculino, predominantemente na faixa etária dos 30 anos, estado civil solteiro, com formação em ciências da vida e que integram o departamento de Pesquisa e Desenvolvimento. Além disso, observou-se que os colaboradores que residem a uma distância inferior a 10 km do local de trabalho apresentam uma probabilidade mais elevada de deixar a empresa em comparação com aqueles que residem a 20 ou 30 km.

Esses insights sugerem que a retenção de colaboradores do sexo masculino na faixa etária dos 30 anos, solteiros, com formação em ciências da vida e que atuam no departamento de Pesquisa e Desenvolvimento pode ser uma área crítica a ser abordada para melhorar a permanência desses profissionais na empresa. Além disso, estratégias de engajamento ou benefícios específicos podem ser implementados para mitigar o atrito identificado. Quanto à proximidade geográfica, medidas para fortalecer a relação e o comprometimento dos colaboradores que residem a menos de 10 km podem ser consideradas, visando reduzir a probabilidade de saída nesse grupo específico.

2. Experiência do colaborador



Ao analisarmos os dados relativos à experiência dos colaboradores, evidencia-se que a probabilidade de um colaborador deixar a empresa é mais significativa nos primeiros anos, embora essa tendência possa persistir até o décimo ano de serviço. Notavelmente, a falta de progresso na carreira ao longo de um período de cinco anos emerge como um fator crítico, aumentando a propensão de um colaborador para deixar a organização.

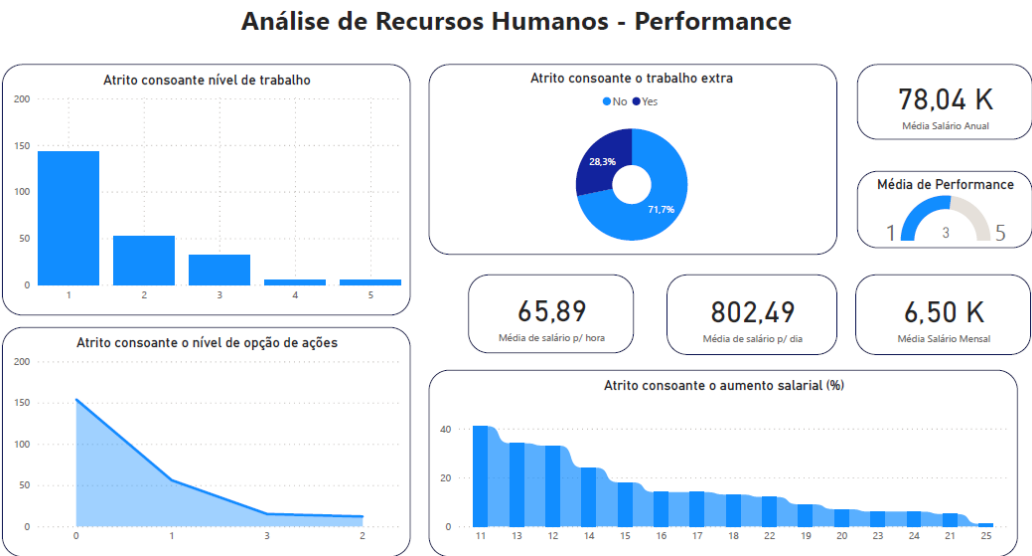
Quanto ao histórico de empregos, destaca-se que a maioria dos colaboradores que trabalhou em apenas uma empresa é mais propensa a sair. Esse padrão inesperado pode sugerir a necessidade de examinar mais profundamente as razões subjacentes a essa correlação aparente, considerando variáveis como insatisfação, falta de oportunidades de crescimento ou outros fatores específicos.

Surpreendentemente, constatou-se que pessoas com uma frequência de viagens mais baixa tendem a apresentar maior atrito em comparação com aquelas que não viajam com frequência. Esta descoberta inusitada pode indicar que a falta de diversidade nas experiências profissionais ou a ausência de desafios relacionados a viagens podem influenciar negativamente o comprometimento dos colaboradores.

Além disso, os dados revelam que os colaboradores que deixaram a empresa muitas vezes enfrentavam desafios no equilíbrio entre vida pessoal e profissional, indicando a importância de estratégias que promovam um ambiente de trabalho saudável e que incentivem o equilíbrio entre as responsabilidades profissionais e pessoais.

Em resumo, as conclusões apontam para a importância de políticas de progresso na carreira, considerações mais aprofundadas sobre os motivos subjacentes ao atrito de colaboradores com experiência em apenas uma empresa, a compreensão dos impactos das viagens no engajamento dos colaboradores e a implementação de medidas para promover um equilíbrio saudável entre vida pessoal e profissional.

3. A Performance dos colaboradores



Ao analisarmos o desempenho dos colaboradores, constata-se que muitos apresentaram uma performance medíocre, evidenciada por níveis de trabalho relativamente baixos, e uma significativa parcela não realizava horas extras. Esta tendência de desempenho reflete-se também nos salários, onde, apesar de um salário anual bruto em torno de 78 mil euros, os colaboradores com menor volume

de opções de ações e um aumento salarial inferior a 15% demonstraram uma propensão maior a deixar a empresa.

É importante relacionar essas observações com o contexto do custo de vida em Portugal. O custo de vida pode variar consideravelmente dependendo da região do país, mas, em geral, Portugal é conhecido por ter um custo de vida mais acessível em comparação com outros países europeus. No entanto, desafios econômicos e diferenças regionais podem impactar a percepção dos colaboradores sobre a adequação de seus salários às despesas diárias.

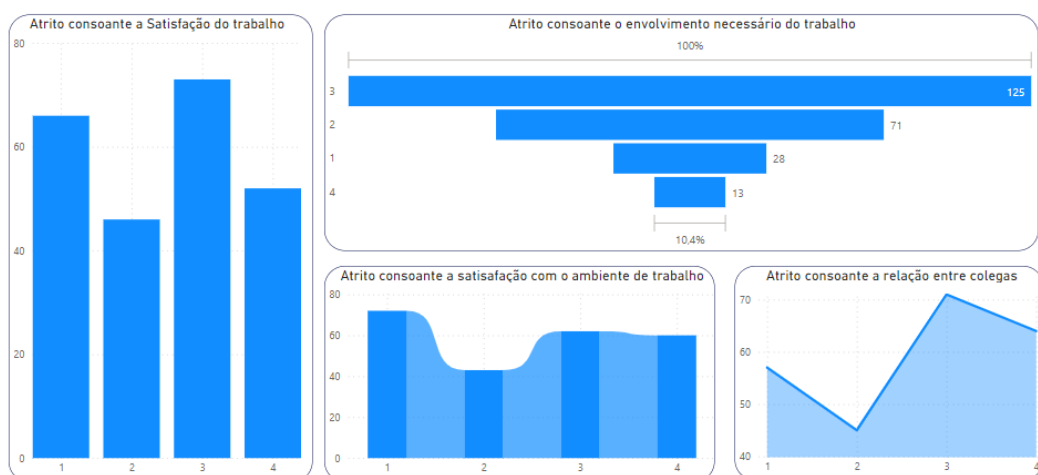
A relação entre desempenho e atrito sugere a necessidade de uma revisão das práticas de gestão de desempenho e compensação. Investir em programas de reconhecimento, oferecer oportunidades de desenvolvimento profissional e garantir aumentos salariais competitivos podem ser estratégias eficazes para melhorar o engajamento e a retenção dos colaboradores.

Ademais, considerando o ambiente econômico e o custo de vida em Portugal, é crucial que as políticas de remuneração estejam alinhadas com as expectativas e necessidades dos colaboradores. A oferta de benefícios que visam atender às demandas específicas do custo de vida em diferentes regiões pode contribuir significativamente para a satisfação e retenção da equipe.

Assim, a análise da relação entre desempenho, remuneração e atrito oferece insights valiosos para o desenvolvimento de estratégias personalizadas, visando aprimorar o ambiente de trabalho e a estabilidade dos colaboradores, considerando as nuances do contexto econômico em Portugal.

4. A Satisfação do colaborador

Análise de Recursos Humanos - Satisfação



Quanto à satisfação dos colaboradores na empresa, é de suma importância cultivar um ambiente de trabalho positivo, pois isso não apenas mantém a equipe motivada, mas também contribui para o bem-estar geral. Na presente análise, nota-se que não se evidencia um ambiente negativo, o que leva à conclusão de que o atrito observado não está relacionado a um clima organizacional desfavorável, à falta de harmonia com os colegas ou à necessidade de maior envolvimento no trabalho.

A ausência de indicadores que apontem para um mal-estar no ambiente de trabalho sugere que outros fatores, como aspectos de gestão, desenvolvimento profissional, reconhecimento ou benefícios, podem estar influenciando a motivação e a permanência dos colaboradores na empresa.

Assim, ao buscar compreender as razões subjacentes ao atrito, é crucial explorar áreas como práticas de gestão, oportunidades de crescimento profissional, programas de reconhecimento e benefícios oferecidos. Uma abordagem holística, considerando diversos aspectos que impactam a satisfação no trabalho, permitirá desenvolver estratégias mais eficazes para a retenção de talentos e a promoção de um ambiente de trabalho verdadeiramente motivador.

Data Mining e Web Analytics

1. Metodologia

Os dados utilizados nesta análise foram obtidos do Kaggle, originados de um estudo elaborado pela IBM. Vale ressaltar que os dados são fictícios e têm como propósito simular o atrito nas empresas, proporcionando insights para possíveis melhorias.

Na fase inicial, concentramo-nos no tratamento e pré-processamento dos dados. Identificamos e removemos colunas com baixa variabilidade, visando otimizar a qualidade das informações. Dado que a base de dados não estava totalmente compatível com os modelos que planejávamos utilizar, foi crucial realizar a categorização dos dados. Esse processo foi conduzido com o objetivo de posteriormente transformá-los em variáveis binárias (0 e 1).

Posteriormente, organizamos os dados em diferentes conjuntos, simplificando sua análise futura com informações semelhantes. Essas tabelas foram agrupadas com base em critérios como demografia, experiência, satisfação e desempenho. Com a conclusão dessa etapa, avançamos para a fase de análise.

Para a Análise Descritiva, optamos por explorar as regras de associação, buscando compreender quais fatores predominam em um indivíduo que o levam a sair da empresa (Atrito-Sim). Já para a Análise Preditiva, escolhemos implementar o modelo de Random Forest ou Floresta Aleatória. Esta escolha foi motivada pela sua capacidade de proporcionar uma análise mais aprofundada e resultados mais precisos.

2. Pré-Processamento e Tratamento de Dados

Os dados foram inicialmente carregados a partir de um arquivo CSV contendo informações sobre a gestão de Recursos Humanos. Uma análise descritiva foi realizada para entender melhor a natureza dos dados. Algumas colunas consideradas irrelevantes para a análise foram removidas, tais como 'StandardHours', 'Over18', 'EmployeeCount', 'EmployeeNumber' e 'PerformanceRating'.

Foram criados mapas para transformar valores numéricos em categorias mais significativas. Esses mapas foram aplicados a colunas relevantes, como 'EnvironmentSatisfaction', 'Education', 'JobLevel' e 'StockOptionLevel'. Variáveis contínuas, como 'Age', 'DistanceFromHome', 'DailyRate', 'HourlyRate', 'MonthlyRate' e outras, foram categorizadas em intervalos específicos. Isso proporcionou uma compreensão mais clara e facilitou a análise dessas variáveis.

Com o intuito de facilitar a análise, os dados foram agrupados em conjuntos específicos:

- Demografia: Informações demográficas, como idade, distância do trabalho, educação, gênero e estado civil.
- Performance: Indicadores de desempenho, incluindo taxas diárias, níveis de emprego, salários, entre outros.
- Satisfação: Medidas de satisfação no trabalho, como satisfação ambiental, envolvimento no trabalho e satisfação no trabalho.
- Experiência: Informações relacionadas à experiência profissional, como viagens de negócios, departamento, cargo, entre outras.

Foram criados datasets específicos para cada conjunto mencionado acima, mantendo as variáveis categóricas em sua forma original. Além disso, versões

dummy (binárias) desses datasets foram geradas usando a função `get_dummies` do `pandas`.

Essas etapas de pré-processamento e criação de datasets visam tornar os dados mais acessíveis e preparados para análises descritivas e preditivas no contexto do projeto de otimização de estratégias de retenção de recursos humanos., como modelagem preditiva ou análise exploratória. Cada transformação foi cuidadosamente documentada para garantir transparência e reprodutibilidade.

Análise Descritiva

A Análise Descritiva realizada empregou o modelo de Regras de Associação para aprofundar a compreensão dos fatores que desencadeiam a rotatividade dentro da organização. O código abaixo apresenta a aplicação dessa abordagem no conjunto de dados relacionado à Demografia:

```
frequent_itemsets = apriori(demo_data_dummy, min_support=0.05, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
attrition_yes_rules = rules[rules['consequents'].apply(lambda x: any(item in x for item in ['Attrition_Yes']))]

print("Demografia - Regras de Associação onde o Atrito positivo é consequência:")
print(attrition_yes_rules)
```

Python

É relevante observar que foi estabelecido um limite máximo para a métrica "Lift", que indica a relação entre o antecedente e o consequente. Posteriormente, realizou-se uma filtragem para exibir apenas os resultados nos quais o consequente é "Attrition_Yes", ou seja, quando o atrito é positivo.

Essa mesma abordagem foi aplicada aos demais conjuntos de dados, com exceção do conjunto relacionado à Satisfação, onde a métrica utilizada foi "confiança" em vez de "lift":

```
frequent_itemsets = apriori(sat_data_dummy, min_support=0.05, use_colnames=True)
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.1)
attrition_yes_rules = rules[rules['consequents'].apply(lambda x: any(item in x for item in ['Attrition_Yes']))]

print("Satisfação - Regras de Associação onde o Atrito positivo é consequência:")
print(attrition_yes_rules)
```

Python

Os resultados obtidos foram os seguintes:

Para Demografia:

Education_Bachelor -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 17.31%• Lift: 1.07 Funcionários com bacharelado têm uma probabilidade ligeiramente superior de atrito em comparação com a média.
Gender_Male -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 17.01%• Lift: 1.05 Homens têm uma probabilidade ligeiramente superior de atrito em comparação com a média.
MaritalStatus_Single -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 25.53%• Lift: 1.58 Interpretação: Funcionários solteiros têm uma probabilidade significativamente maior de atrito em comparação com a média.
Gender_Male, Age_Middle Age -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 11.34%• Lift: 1.05 Homens de meia-idade têm uma probabilidade ligeiramente maior de atrito em comparação com a média.
MaritalStatus_Single, Age_Middle Age -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 22.99%• Lift: 1.43 Funcionários solteiros de meia-idade têm uma probabilidade significativamente maior de atrito em comparação com a média.
MaritalStatus_Single, DistanceFromHome_Close -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 23.17%• Lift: 1.44 Funcionários solteiros que moram perto têm uma probabilidade significativamente maior de atrito em comparação com a média.
Gender_Male, DistanceFromHome_Close -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 10.77%• Lift: 1.10 Homens que moram perto têm uma probabilidade ligeiramente maior de atrito em comparação com a média.
MaritalStatus_Single, DistanceFromHome_Close -> Attrition_Yes <ul style="list-style-type: none">• Confiança: 23.17%• Lift: 1.44 Funcionários solteiros que moram perto têm uma probabilidade significativamente maior de atrito em comparação com a média.

Embora o resultado da confiança não seja elevado, podemos concluir que, em termos demográficos um homem solteiro, de meia-idade, que vive perto do local onde trabalha e com uma licenciatura, seja mais provável de sair de uma empresa.

Para Performance:

<p>DailyRate_Low Rate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 17.17% • Lift: 1.07 <p>Funcionários com uma taxa diária baixa têm uma probabilidade ligeiramente superior de atrito em comparação com a média.</p>
<p>HourlyRate_Medium Rate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.61% • Lift: 1.03 <p>Funcionários com uma taxa horária média têm uma probabilidade ligeiramente superior de atrito em comparação com a média.</p>
<p>JobLevel_Associate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 26.34% • Lift: 1.63 <p>Funcionários com nível associado têm uma probabilidade significativamente maior de atrito em comparação com a média.</p>
<p>MonthlyIncome_Low Income -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 21.77% • Lift: 1.35 <p>Funcionários com renda mensal baixa têm uma probabilidade maior de atrito em comparação com a média.</p>
<p>MonthlyRate_Medium Rate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.85% • Lift: 1.04 <p>Funcionários com uma taxa mensal média têm uma probabilidade ligeiramente superior de atrito em comparação com a média.</p>
<p>DailyRate_Low Rate, JobLevel_Associate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 18.14% • Lift: 1.12 <p>Funcionários com uma taxa diária baixa e nível associado têm uma probabilidade ainda maior de atrito.</p>
<p>MonthlyIncome_Low Income, JobLevel_Associate -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 17.50% • Lift: 1.08 <p>Funcionários com renda mensal baixa e nível associado têm uma probabilidade significativamente maior de atrito.</p>

Neste contexto, funcionários com salários mais baixos, taxas diárias baixas, níveis associados e certas combinações desses fatores têm uma probabilidade maior de atrito. Essas informações podem orientar estratégias de retenção e gestão de recursos humanos, ajudando a empresa a abordar especificamente os grupos de funcionários mais propensos ao atrito.

Na análise de regras de associação, apenas se encontrou uma em que o consequente fosse positivo, e essa foi:

JobInvolvement_Good -> Attrition_Yes

- Confiança: 14.40%
- Lift: 0.89

Funcionários com bom envolvimento no trabalho têm uma probabilidade ligeiramente menor de atrito em comparação com a média.

Esta descoberta sugere que funcionários com um alto nível de satisfação em relação ao envolvimento no trabalho têm uma probabilidade menor de deixar a empresa, indicando uma possível correlação positiva entre a satisfação no trabalho e a retenção de funcionários.

Essa informação pode ser valiosa para a empresa ao implementar estratégias para melhorar a satisfação dos funcionários, especialmente em termos de envolvimento no trabalho. Promover um ambiente de trabalho positivo pode contribuir para a redução do atrito entre os colaboradores.

E por fim, no contexto de Experiência:

Department_Sales -> Attrition_Yes

- Confiança: 20.63%
- Lift: 1.28

Funcionários no departamento de vendas têm uma probabilidade significativamente maior de atrito em comparação com a média.

NumCompaniesWorked_<10_years -> Attrition_Yes

- Confiança: 16.12%
- Lift: 1.00

Funcionários que trabalharam em menos de 10 empresas têm uma probabilidade média de atrito.

TotalWorkingYears_<10_years -> Attrition_Yes

- Confiança: 19.72%
- Lift: 1.22

Funcionários com menos de 10 anos de experiência total têm uma probabilidade maior de atrito.

<p>TrainingTimesLastYear_<10_years -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.12% • Lift: 1.00 <p>Funcionários que receberam treinamento menos de 10 vezes no último ano têm uma probabilidade média de atrito.</p>
<p>YearsAtCompany_<10_years -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 17.73% • Lift: 1.10 <p>Funcionários com menos de 10 anos na empresa têm uma probabilidade ligeiramente maior de atrito.</p>
<p>YearsInCurrentRole_<10_years -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.67% • Lift: 1.03 <p>Funcionários com menos de 10 anos na função atual têm uma probabilidade média de atrito.</p>
<p>YearsSinceLastPromotion_<10_years -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.31% • Lift: 1.01 <p>Funcionários com menos de 10 anos desde a última promoção têm uma probabilidade média de atrito.</p>
<p>YearsWithCurrManager_<10_years -> Attrition_Yes</p> <ul style="list-style-type: none"> • Confiança: 16.75% • Lift: 1.04 <p>Funcionários com menos de 10 anos com o atual gerente têm uma probabilidade média de atrito.</p>

Os funcionários em certos cenários, como aqueles no departamento de vendas, com menos de 10 anos de experiência total, ou com menos de 10 anos na empresa, têm uma probabilidade maior de atrito.

Análise Preditiva

O Random Forest foi escolhido para a análise preditiva devido à sua capacidade comprovada de oferecer alta precisão, tratamento eficiente de dados categóricos e robustez a valores ausentes, tornando-o uma escolha versátil e confiável.

```
#Separando variável resposta e variáveis features
X = demo_data_dummy.drop(['Attrition_Yes', 'Attrition_No'], axis = 1)
y = demo_data_dummy['Attrition_Yes']

# Escalonando os dados
sc = StandardScaler()
X = sc.fit_transform(X)

#Dividindo Treino e Teste
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)

# Normalizando os dados
X_train = normalize(X_train, norm='l2', axis=1, copy=True, return_norm=False)
X_test = normalize(X_test, norm='l2', axis=1, copy=True, return_norm=False)

rfc = RandomForestClassifier(n_estimators=200)
rfc.fit(X_train, y_train)
pred_rfc = rfc.predict(X_test)
#print(rfc.score(X_test, y_test))
print('Dados relevantes na Demografia:')
for i in range (len(demo_data_dummy.columns[2:])):
    if rfc.feature_importances_[i]>0.05:
        print(demo_data_dummy.columns[2:][i])
        print(rfc.feature_importances_[i])
print()
# Cria a matriz de confusão
cm = confusion_matrix(y_test, pred_rfc)

# Adiciona labels à matriz de confusão
classes = ['Yes', 'No']
cm_df = pd.DataFrame(cm, index=classes, columns=classes)

# Exibe a matriz de confusão
print('Matriz de confusão')
print(cm_df)
print()
print('Relatório de Classificação')
print(classification_report(y_test, pred_rfc))
print()
print("Random Forest - Demografia:", rfc.score(X_test, y_test) )
```

Python

Neste código, conduzimos uma análise preditiva utilizando o algoritmo Random Forest para avaliar a influência de variáveis demográficas na rotatividade.

Os passos incluem a preparação dos dados, com separação de variáveis resposta e features, seguida pelo escalonamento e divisão em conjuntos de treino e teste. Após a normalização, o modelo Random Forest, composto por 200 estimadores, é treinado com os dados de treino. Avaliamos a importância das features, destacando aquelas que têm um impacto significativo na previsão da rotatividade.

Com esta análise vamos poder analisar os Dados Relevantes a cada tabela, onde as características consideradas importantes pelo modelo de Random Forest ao tomar decisões; A matriz de confusão, em que se compara as previsões do modelo com os resultados reais, dividindo os resultados em quatro partes: Verdadeiros Positivos (Previsões corretas de casos positivos (sim)), Falsos Positivos (Previsões incorretas de casos positivos), Verdadeiros Negativos (Previsões corretas de casos

negativos (não)) e Falsos Negativos (Previsões incorretas de casos negativos). Esses números ajudam a avaliar o desempenho do modelo, mostrando onde ele acertou e errou em suas previsões; e por fim o Relatório de Classificação, que incorpora um resumo estatístico que fornece métricas de desempenho do modelo, como precisão (Indica a proporção de previsões corretas para uma determinada classe) , recall (Mostra a capacidade do modelo em identificar todos os casos positivos), F1-Score (Combina precisão e revocação em uma única métrica) e acurácia (Representa a proporção total de previsões corretas do modelo)

Ao aplicar a função `train_test_split` nos conjuntos de dados, o parâmetro `test_size` foi estrategicamente ajustado em diferentes valores. Essa modificação teve como objetivo otimizar o desempenho do classificador Random Forest nos diversos conjuntos de dados. A variação do `test_size` foi uma decisão prática, buscando encontrar um equilíbrio que permitisse uma avaliação robusta do modelo sem comprometer a quantidade de dados disponíveis para treinamento. Assim, por meio da exploração desses diferentes tamanhos de conjuntos de teste, procurou-se atingir resultados mais eficazes na predição de atrito positivo em cada contexto específico.

Foi utilizado código acima para os outros datasets, embora tenha-se mudado o valor do “`test_size`”, para melhores resultados, como mostram os resultados obtidos:

Demografia:

Dados Relevantes na Demografia:

1. Age_Young Adults: 0.111
2. DistanceFromHome_Close: 0.070
3. DistanceFromHome_Far: 0.066
4. DistanceFromHome_Medium: 0.059
5. Education_Bachelor: 0.052
6. Education_Doctor: 0.052
7. Gender_Female: 0.052
8. Gender_Male: 0.056
9. MaritalStatus_Divorced: 0.059
10. MaritalStatus_Married: 0.066
11. MaritalStatus_Single: 0.126

Essas são as características demográficas mais relevantes, ordenadas pela importância atribuída pelo modelo de Random Forest com o novo tamanho do conjunto de teste.

Matriz de Confusão:

	Yes	No
Yes	233	11
No	44	6

- True Positives (TP): 233
- False Positives (FP): 11
- True Negatives (TN): 6
- False Negatives (FN): 44

Relatório de Classificação:					
	precision	recall	f1-score	support	
	0	0.84	0.95	0.89	244
	1	0.35	0.12	0.18	50
accuracy				0.81	294
macro avg	0.60	0.54	0.54		294
weighted avg	0.76	0.81	0.77		294
<ul style="list-style-type: none">Precisão (Precision): A precisão para a classe "0" (Não atrito) é de 84%, enquanto para a classe "1" (Atrito) é de 35%.Revocação (Recall): A revocação para a classe "0" é de 95%, indicando que o modelo identifica corretamente a maioria dos casos "Não atrito". No entanto, a revocação para a classe "1" é baixa, indicando que o modelo tem dificuldade em identificar casos de atrito.F1-Score: O F1-Score, que combina precisão e revocação, é de 0.89 para a classe "0" e 0.18 para a classe "1".Acurácia: A acurácia geral do modelo é de 81%.					

Performance

Dados Relevantes na Performance:					
	1. JobLevel_Associate:	0.055			
	2. OverTime_No:	0.083			
	3. OverTime_Yes:	0.078			
	4. StockOptionLevel_Level 0:	0.052			
Essas são as características de performance mais relevantes, ordenadas pela importância atribuída pelo modelo de Random Forest.					
Matriz de Confusão:					
	Yes	No			
Yes	468	32			
No	69	19			
<ul style="list-style-type: none">True Positives (TP): 468False Positives (FP): 32True Negatives (TN): 19False Negatives (FN): 69					
Relatório de Classificação:					
	precision	recall	f1-score	support	
	0	0.87	0.94	0.90	500
	1	0.37	0.22	0.27	88
accuracy				0.83	588
macro avg	0.62	0.58	0.59		588
weighted avg	0.80	0.83	0.81		588
<ul style="list-style-type: none">Precisão (Precision): A precisão para a classe "0" (Desempenho Adequado) é de 87%, enquanto para a classe "1" (Desempenho Insatisfatório) é de 37%.Revocação (Recall): A revocação para a classe "0" é de 94%, indicando que o modelo identifica corretamente a maioria dos casos de desempenho adequado. No entanto, a revocação para a classe "1" é mais baixa, indicando que o modelo tem dificuldade em identificar casos de desempenho insatisfatório.F1-Score: O F1-Score, que combina precisão e revocação, é de 0.90 para a classe "0" e 0.27 para a classe "1".Acurácia: A acurácia geral do modelo é de 83%.					

Satisfação

Experiência

Dados Relevantes na Experiência:

1. BusinessTravel_Travel_Frequently: 0.055
2. JobRole_Sales Representative: 0.059
3. WorkLifeBalance_Low: 0.062

Essas são as características de experiência mais relevantes, ordenadas pela importância atribuída pelo modelo de Random Forest.

Matriz de Confusão:

	Yes	No
Yes	371	6
No	59	5

- True Positives (TP): 371
- False Positives (FP): 6
- True Negatives (TN): 5
- False Negatives (FN): 59

Relatório de Classificação:				
	precision	recall	f1-score	support
0	0.86	0.98	0.92	377
1	0.45	0.08	0.13	64
accuracy			0.85	441
macro avg	0.66	0.53	0.53	441
weighted avg	0.80	0.85	0.81	441

- **Precisão (Precision):** A precisão para a classe "0" (Experiência Adequada) é de 86%, enquanto para a classe "1" (Experiência Insatisfatória) é de 45%.
- **Revocação (Recall):** A revocação para a classe "0" é de 98%, indicando que o modelo identifica corretamente a maioria dos casos de experiência adequada. No entanto, a revocação para a classe "1" é baixa, indicando que o modelo tem dificuldade em identificar casos de experiência insatisfatória.
- **F1-Score:** O F1-Score, que combina precisão e revocação, é de 0.92 para a classe "0" e 0.13 para a classe "1".
- **Acurácia:** A acurácia geral do modelo é de 85%.

Ao comparar os resultados obtidos através das regras de associação e do modelo de Random Forest para analisar o atrito positivo na empresa, observamos que ambos os métodos proporcionam insights valiosos, embora com abordagens distintas.

Regras de Associação:

- Destaca relações específicas entre variáveis demográficas, financeiras, de satisfação e de experiência e sua influência no atrito positivo.
- Fornece uma visão direta das associações e das probabilidades envolvidas.
- Traz à tona características específicas, como estado civil, gênero, salários, satisfação no trabalho e experiência, que podem estar associadas ao atrito.

Random Forest:

- Utiliza um modelo preditivo abrangente que considera múltiplas características simultaneamente, atribuindo importâncias relativas a cada uma delas.
- Oferece métricas de desempenho, como precisão, recall e F1-score, proporcionando uma avaliação mais quantitativa da eficácia do modelo.
- Indica a relevância de características específicas em diferentes contextos, como demografia, desempenho, satisfação e experiência.

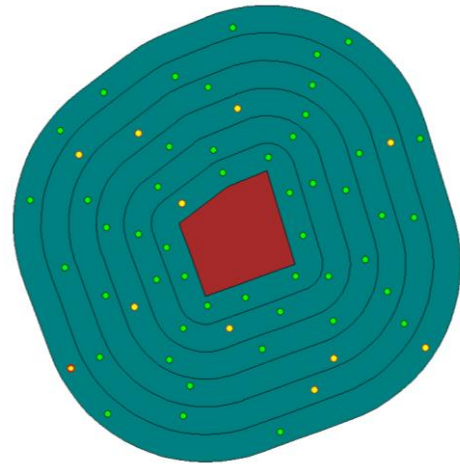
Conclusão Final: Ambos os métodos convergem em pontos essenciais, destacando a importância de fatores como estado civil, salários, satisfação no trabalho e experiência na influência do atrito positivo. Enquanto as regras de associação oferecem interpretações mais diretas e probabilísticas, o Random Forest fornece uma abordagem mais abrangente, com métricas de avaliação de desempenho.

Para a empresa, a combinação dessas abordagens pode ser poderosa. As regras de associação oferecem insights detalhados sobre relações específicas, enquanto o Random Forest fornece uma visão global e métricas que orientam na avaliação da eficácia preditiva do modelo.

Em resumo, ao integrar os resultados dessas abordagens, a empresa pode criar estratégias de retenção mais robustas, abordando de maneira eficaz os fatores que contribuem para o atrito positivo em diferentes aspectos do ambiente de trabalho.

Sistemas de Informação Geográfica

O "buffer" (zona azul) estabelecido em torno das instalações da empresa (delimitada pela área em vermelho) foi meticulosamente projetado para oferecer uma representação visual precisa das flutuações no atrito em distâncias variadas. Este esquema foi concebido de modo a replicar fielmente as características presentes na base de dados, proporcionando uma visão simplificada e de fácil interpretação.



Ao analisar a representação gráfica, nota-se uma redução no número de pessoas que deixam o local, como indicado pelos pontos amarelos. Essa diminuição é particularmente evidente, variando de 1 a 3 pessoas em diferentes zonas do "buffer". A segmentação do "buffer" em seis zonas distintas, cobrindo distâncias que vão de 0 a 30 quilômetros da empresa, oferece uma visão abrangente da influência do espaço geográfico no padrão de saída.

Esses resultados sugerem que a proximidade geográfica exerce um não impacto significativo na mobilidade dos indivíduos em relação à empresa. A variação observada nas diferentes zonas do "buffer" pode indicar diferentes níveis de atrito ou motivações para deixar a empresa.

Economia

A alta rotatividade de funcionários em uma empresa pode desencadear uma série de efeitos prejudiciais que reverberam profundamente em sua saúde econômica. Esta questão complexa transcende a mera flutuação de pessoal e se insere em diferentes áreas da operação empresarial, com implicações diretas nos resultados financeiros e na imagem corporativa. Neste contexto, é crucial analisar os impactos econômicos desse fenômeno e compreender sua relação com a teoria utilitarista de economia.

1. Menos Pessoas a Trabalhar, Menos Produtividade:

Uma taxa elevada de rotatividade de funcionários resulta em uma força de trabalho reduzida. Com menos mãos para realizar as tarefas necessárias, a produtividade da empresa é comprometida. Setores específicos podem ser desfavorecidos, enfrentando carências de competências e conhecimentos, o que prejudica ainda mais a eficiência operacional.

2. Menos Lucro e Possibilidade de Baixar Salários:

A redução da produtividade, resultante da alta rotatividade, impacta diretamente nos lucros da empresa. A instabilidade causada pela entrada e saída contínuas de funcionários pode levar a perdas financeiras significativas. Além disso, a pressão para manter os lucros pode levar a medidas extremas, como a redução de salários, afetando a motivação e satisfação dos colaboradores remanescentes.

3. Custos de Recrutamento, Contratação e Demissão:

O processo de recrutar, contratar e demitir funcionários é oneroso. A constante necessidade de preencher posições vagas impõe custos significativos à empresa, incluindo despesas com anúncios de emprego, entrevistas, treinamento e despesas administrativas relacionadas à contratação e demissão. Esses custos adicionais impactam diretamente a margem de lucro da empresa.

4. Destruição da Imagem Corporativa:

A alta rotatividade pode minar a reputação da empresa, prejudicando sua imagem corporativa. Os clientes percebem a instabilidade interna e podem ficar desconfiados em relação à qualidade dos produtos ou serviços oferecidos. A falta de estabilidade também desincentiva a fidelidade do cliente, levando-os a buscar opções mais estáveis e confiáveis no mercado.

5. Relação com a Teoria Utilitarista de Economia:

A teoria utilitarista enfatiza a maximização da felicidade ou utilidade geral. No contexto empresarial, a alta rotatividade pode ser vista como contraproducente, pois reduz a satisfação dos funcionários, clientes e outros stakeholders. Maximizar a utilidade geral significa criar um ambiente de trabalho estável e próspero, promovendo a satisfação e a eficiência para todos os envolvidos.

Em conclusão, a alta rotatividade de funcionários não é apenas uma questão de recursos humanos, mas tem implicações profundas no desempenho económico global de uma empresa. A abordagem utilitarista destaca a importância de criar condições favoráveis para todas as partes interessadas, sugerindo que a estabilidade e a satisfação geral são fundamentais para o sucesso sustentável nos negócios.

Referências

Subhash, P. (2018). IBM HR Analytics - Attrition Dataset. Kaggle. Recuperado de: <https://www.kaggle.com/datasets/pavansubhasht/ibm-hr-analytics-attrition-dataset>

INSEAD Data Analytics. (2018). IBM Attrition VSS. Recuperado de: [https://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/IBM Attrition VSS.html](https://inseaddataanalytics.github.io/INSEADAnalytics/groupprojects/January2018FBL/IBM%20Attrition%20VSS.html)

Romeira, A. (2022, 19 de fevereiro). Taxa de rotatividade estabilizou, mas disparidades salariais permanecem nas empresas de recursos humanos. Jornal Económico. <https://jornaleconomico.sapo.pt/noticias/taxa-de-rotatividade-estabilizou-mas-disparidades-salariais-permanecem-nas-empresas-de-recursos-humanos-850520/>

Redação. (2023, 27 de janeiro). As 12 principais causas dos pedidos de demissão. Você RH. <https://vocerh.abril.com.br/futurodotrabalho/as-12-principais-causas-dos-pedidos-de-demissao>

DOT Digital Group. (2023, 29 de maio). Entenda a relação entre burnout e turnover. LinkedIn. <https://www.linkedin.com/pulse/entenda-rela%C3%A7%C3%A3o-entre-burnout-e-turnover-dot-digital-group/?originalSubdomain=pt>

Mercer. (2023, September 13). Empresas revelam dificuldades na retenção de talento. Mercer. <https://www.mercer.com/pt-pt/about/newsroom/empresas-revelam-dificuldades-na-retencao-de-talento/>

Esolidar. (2020, February 11). Quanto custa um funcionário desligado e turnover elevado para a empresa? Impacto Social. <https://impactosocial.esolidar.com/pt-pt/2020/02/11/quanto-custa-um-funcionario-desligado-e-turnover-elevado-para-a-empresa/>

Vangardi. (2020, July 21). Turnover: O que é e como calcular a rotatividade de funcionários? Vangardi. <https://vangardi.com.br/turnover/>

Gianotto, D. (2022, January 3). Turnover: O que é, tipos e impactos para a empresa. LinkedIn. <https://www.linkedin.com/pulse/turnover-o-que-%C3%A9-tipos-e-impactos-para-empresa-debora-gianotto/?originalSubdomain=pt>

MarqPontoApp. (2022, November 23). O que é rotatividade e qual seu impacto para a empresa? LinkedIn. <https://www.linkedin.com/pulse/o-que-%C3%A9-rotatividade-e-qual-seu-impacto-para-empresa-marqpontoapp/?trk=pulse-article&originalSubdomain=pt>