**Machine Learning – Não Supervisionado**

Este documento é apenas um breve relatório do trabalho feito. Os algoritmos, raciocínio, desenvolvimento do projeto foi todo feito em Jupyter Notebook, Python. Que também contem boa parte destas informações.

As principais bibliotecas utilizadas foram: Pandas, Numpy, Seaborn, Matplotlib, Counter, StadardScaler, Kmeans, Silhouette\_Score, PCA, Axes3D.

**RESULTADO:**  
A segmentação dos funcionários em clusters com base em seus atributos e comportamentos é um passo fundamental para desenvolver estratégias de retenção mais eficazes.

Com base no trabalho de Machine Learning, não supervisionado, utilizando K-Means e PCA, definimos uma divisão nos dados em seis partes (Clusters):

* Cluster 0: homens que gostam de seus empregos
* Cluster 1: mulheres iniciando a carreira
* Cluster 2: profissionais de alto desempenho
* Cluster 3: homens que não gostam de seus empregos
* Cluster 4: funcionários seniores
* Cluster 5: pessoas que fazem longos trajetos

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Encontramos a seguintes taxas de desligamento nos clusters:

|  |  |
| --- | --- |
| **Cluster** | **Valor** |
| pessoas que fazem longos trajetos | 21,83% |
| homens que não gostam de seus empregos | 18,91% |
| profissionais de alto desempenho | 18,63% |
| homens que gostam de seus empregos | 16,56% |
| mulheres iniciando a carreira | 15,67% |
| funcionários seniores | 7,31% |

Considerando os Clusters, a taxa de desligamento que encontramos apresentados e as melhores práticas de gestão de pessoas:

**Pessoas que Fazem Longos Trajetos**

* **Opções Remotas:** Ampliar a flexibilidade para trabalho remoto, mesmo que seja em alguns dias da semana, pode ser um grande atrativo para reduzir o desgaste causado por longas jornadas.
* **Incentivos de Transporte:** Oferecer benefícios como vale-transporte com valores mais altos, reembolso de pedágios ou até mesmo um programa de carpooling podem aliviar o custo financeiro dos deslocamentos.
* **Estações de Trabalho Confortáveis:** Garantir que o ambiente de trabalho seja ergonômico e confortável pode compensar o tempo gasto no trânsito.
* **Programas de Bem-estar:** Oferecer programas de ginástica laboral, meditação ou outras atividades que ajudem a reduzir o estresse causado pelos longos trajetos.

**Homens que Não Gostam de Seus Empregos**

* **Entrevistas Individuais:** Realizar conversas individuais com os gerentes para entender as causas da insatisfação e desenvolver planos de ação personalizados.
* **Oportunidades de Desenvolvimento:** Oferecer programas de treinamento e desenvolvimento que permitam aos funcionários adquirirem novas habilidades e crescer profissionalmente dentro da empresa.
* **Reconhecimento:** Implementar programas de reconhecimento para valorizar as contribuições dos funcionários e aumentar o engajamento.
* **Rotação de Funções:** Permitir que os funcionários experimentem diferentes funções pode renovar o interesse e a motivação.

**Profissionais de Alto Desempenho**

* **Planos de Carreira Individualizados:** Desenvolver planos de carreira personalizados que alinhem os objetivos dos funcionários com os objetivos da empresa.
* **Oportunidades de Liderança:** Delegar responsabilidades e oferecer oportunidades de liderança podem aumentar o senso de propósito e a satisfação profissional.
* **Mentoria:** Conectar os profissionais de alto desempenho com mentores experientes pode acelerar o desenvolvimento de suas carreiras.
* **Recompensas Financeiras e Não Financeiras:** Oferecer remuneração competitiva, bonificações e outros benefícios atrativos pode ajudar a reter esses talentos.

**Homens que Gostam de Seus Empregos**

* **Fortalecer a Cultura:** Investir em ações que fortaleçam a cultura organizacional e criem um ambiente de trabalho positivo e colaborativo.
* **Programas de Bem-estar:** Oferecer programas de bem-estar que promovam a saúde física e mental dos funcionários.
* **Reconhecimento:** Implementar programas de reconhecimento que valorizem as contribuições dos funcionários e aumentem o engajamento.
* **Oportunidades de Networking:** Promover eventos e atividades que permitam aos funcionários se conectarem e construir relacionamentos.

**Mulheres Iniciando a Carreira**

* **Mentoria e Coaching:** Oferecer programas de mentoria e coaching para auxiliar no desenvolvimento profissional e na adaptação à cultura da empresa.
* **Redes de Apoio:** Criar redes de apoio para mulheres, como grupos de networking e programas de desenvolvimento de liderança.
* **Flexibilidade:** Oferecer opções de trabalho flexíveis, como horários ajustáveis e trabalho remoto, para facilitar a conciliação entre vida pessoal e profissional.

**Funcionários Seniores**

* **Reconhecimento:** Valorizar a experiência e o conhecimento dos funcionários seniores através de programas de reconhecimento e celebração de suas contribuições.
* **Mentoria Reversa:** Incentivar a mentoria reversa, onde os funcionários seniores podem compartilhar seus conhecimentos com os mais jovens.
* **Transição para a Aposentadoria:** Oferecer programas de transição para a aposentadoria que preparem os funcionários para essa nova fase de vida.

**Recomendações Adicionais:**

* **Pesquisa de Clima Organizacional:** Realizar pesquisas de clima organizacional regularmente para identificar as principais causas de insatisfação e tomar medidas corretivas.
* **Análise de Dados:** Utilizar dados para identificar padrões e tendências que possam influenciar a rotatividade.
* **Personalização:** Desenvolver estratégias de retenção personalizadas para cada segmento de funcionários.

**OBS:** As estratégias de retenção devem ser adaptadas à cultura e aos objetivos da sua empresa. É importante monitorar os resultados das iniciativas implementadas e fazer ajustes conforme necessário.

**Problema e Solução:**

Uma determinada com problemas de retenção de funcionários nos solicitou um trabalho descobrir os motivos de Turn over da empresa e encontrar alternativas para aumentar esta retenção. Para fim deste trabalho, já que o estamos apresentando abertamente, chamaremos a empresa de “CIA-A”.

Por questões obvias retiramos os nomes dos funcionários, modificamos os nomes dos departamentos, valores de salários, fizemos pesquisa de satisfação e chegamos a uma determinada planilha.

Com acesso aos dados de funcionários da “CIA-A”, estão incluídas informações demográficas, histórico de desempenho, tempo de serviço na empresa, rotatividade...

Nosso **Objetivo** e entender melhor os tipos de funcionários buscando desenvolver medidas baseadas em dados para a retenção dos funcionários dentro de cada um dos segmentos que vamos encontrar dentro da empresa.

Para esta tarefa usaremos ferramentas de  **\*\*ML não supervisionadas\*\***, para segmentar os funcionários e fazer recomendações para aumentar retenção.

1.  **Preparação de dados & Análise Exploratória**

**Entendendo os dados:**

Base de dados formecida pela empresa consta as seguintes informações:

* funcionario\_id = codigo do funcionário
* idade = idade do funcionário
* sexo = sexo do funcionário
* distancia\_casa = distancia de casa
* nivel\_cargo = nivel do cargo dentro da empresa
* departamento = área em que o funcionário trabalha dentro da empresa
* salario\_mensal = valor so salário mensal do funcionário
* avaliacao\_performance = pontuação atribuido pelo resultado do trabalho
* satisfacao\_trabalho = nota que o funionário atribui a satisfação com o trabalho
* desligamento = se o funcionário saiu da empresa

Com 1480 funcionários (ativos(1239) ou não(241)) 887 do sexo masculino e 593 do sexo feminino divididos em 4 departamentos. Produção e montagem (962), Comercial (448), Recursos Humanos (65), TI (5).

Encontramos uma taxa de rotatividade de 16,28% com base em um agrupamentento realizado nos dados encontramos que as pessoas que tendem a ficar na empresa são as pessoas com mais idade, que moram perto, que trabalham na produção.

**Análise Geral dos Dados:**

* Idade e gênero parecem estar distribuídos de forma bastante uniforme
* As pessoas tendem a morar mais perto do trabalho
* Nível de trabalho e renda estão correlacionados
* Há menos pessoas de alto desempenho
* A maioria das pessoas está feliz com os empregos
* Há poucas pessoas em RH e TI em comparação com os outros departamentos

**2. Agrupamento de K-Means – 1ª tecnica**

o **StandardScaler** vai garantir que todas as variáveis tenham uma média de 0 e desvio padrão de 1, o que é especialmente útil para o K-means, já que ele é sensível à escala das variáveis. Assim, o algoritmo poderá identificar clusters com base nas características padronizadas dos funcionários, levando a uma segmentação mais precisa.

Posteriormente criamos um modelo para determinar qual o numero ideal de clusters esta análise foi feita avaliando a inércia e pontuação de silhueta. Com isso encontramos os gráficos:

|  |  |
| --- | --- |
| Gráfico, Gráfico de linhas  Descrição gerada automaticamente | Gráfico, Gráfico de linhas  Descrição gerada automaticamente |

Com base no ponto de inflexão na curva de inércia, conhecido como o "cotovelo" (elbow) e a pontuação de silhueta para diferentes valores de k onde um valor mais alto da pontuação de silhueta sugere que os clusters são mais bem definidos, decidimos iniciar os testes com 4 clusters.

Treinamos o modelo K-means (Machine Learning não supervisionada) do scikit-learn. Com todos os dados devidamente preparados (ver o arquivo Projeto\_RH\_CIA-A.ipynb)

Plotamos um HEATMAP para análise dos Clusteres:

Calendário

Descrição gerada automaticamente

Chegamos as seguintes classificação dos clusters de acordo com o HeatMap:

\* Cluster 0: funcionários juniores, de Produção e Montagem

\* Cluster 1: funcionários seniores

\* Cluster 2: funcionários do Comercial e TI

\* Cluster 3: funcionários de RH

**3. PCA – 2º Técnica**

Vamos reduzir a dimensionalidade dos dados com o objetivo de encontrar os dois principais componentes que capturam a maior parte da variação dos dados.

Após treinar o modelo de PCA encontramos uma razão de variância baixa, cerca de 38% com dois componentes:

\* Componente 1: idade mais alta, nível cargo, renda mensal = mais sênior

\* Componente 2: menor = Comercial, maior = Produção e Montagem

Encontramos que o agrupamento por departamento tem um grande impacto nos dados e fraciona os modelos como se pode ver abaixo.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

Vamos fazer nova rodada para ver se retirando essa informação do modelo as informações ficam mais homogenias

4. K-Means – 2º Round

Para aprofundar o entendimento dos dados, vamos retirar os departamentos, porque esse criam um separação dos dados, o que atrapalha o entendimento dos doados mais profundamente.

Então com um novo dataset, sem os departamentos novamente buscamos os gráficos de inércia e de “Silhouette Score”.

|  |  |
| --- | --- |
|  |  |
|  |  |

A escolha do número ideal de clusters em modelos não supervisionados é um processo iterativo e, em certa medida, subjetivo. Vamos utilizar diversos critérios e métricas para avaliar a qualidade dos agrupamentos, mas a decisão final vai envolver um julgamento profissional, considerando o contexto do problema e o conhecimento do domínio.

Então foi avaliado três opções:

1. 3 Clusters:

Gráfico

Descrição gerada automaticamente

O que os clusters dizem:

\* Cluster 0: funcionários juniores, sexo feminino

\* Cluster 1: funcionários juniores,  sexo masculino

\* Cluster 2: funcionários seniores

1. 4 Clusters:

Gráfico

Descrição gerada automaticamente com confiança baixa

O que os clusters dizem:

\* Cluster 0: Profissionais Experientes e Satisfeitos

\* Cluster 1: Novos Talentos Femininos

\* Cluster 2: funcionários de alto desempenho

\* Cluster 3: Novos Talentos Masculinos

1. 6 Clusters:

Gráfico

Descrição gerada automaticamente com confiança média

O que os clusters dizem:

\* Cluster 0: homens que gostam de seus empregos

\* Cluster 1: mulheres iniciando a carreira

\* Cluster 2: profissionais de alto desempenho

\* Cluster 3: homens que não gostam de seus empregos

\* Cluster 4: funcionários seniores

\* Cluster 5: pessoas que fazem longos trajetos

Nos parece que afinal a Clusterização com 6 clusters tras uma explicação interessante e robusta o suficiente.

5. PCA – Round 2

A ideia é verificar como vai se comportar o PCA sem os departamentos seguindo o que foi feito com o K-means

A nova razão da variância já explica quase 50% dos dados o que é bem superior. E agora já temos os dados bem mais fuidos no gráfico, sem estar separados como o PCA-round1 vamos comparar:

|  |  |
| --- | --- |
| PCA – Round 1 | PCA – Round 2 |
|  |  |

Como a visualização agora ficou mais complicada resolvemos também fazer uma plotagem dos dados em 3D.

Gráfico, Gráfico de dispersão

Descrição gerada automaticamente

**6. EDA em clusters**

Neste processo rearrumamos os dados, para conseguirmos visualizar os dados frente a questão principal o desligamento.

|  |  |
| --- | --- |
| **Cluster** | **Valor** |
| pessoas que fazem longos trajetos | 21,83% |
| homens que não gostam de seus empregos | 18,91% |
| profissionais de alto desempenho | 18,63% |
| homens que gostam de seus empregos | 16,56% |
| mulheres iniciando a carreira | 15,67% |
| funcionários seniores | 7,31% |