

Relatório da Análise de Dados de Funcionários

Este relatório descreve o processo de análise de dados realizado em três conjuntos de dados relacionados aos funcionários de uma empresa: impacto do trabalho remoto na saúde mental, programas de bem-estar e feedback dos funcionários, e progressão de carreira.

Objetivo:

O objetivo principal da análise foi integrar e limpar os dados para futuras análises e modelagem, garantindo a qualidade e confiabilidade das informações.

Etapas Realizadas:

1. Coleta e Importação dos Dados:

- Três conjuntos de dados foram coletados, cada um abordando um aspecto diferente da experiência dos funcionários.
- Os dados foram importados para o ambiente de análise utilizando a biblioteca Pandas em Python.

2. Preparação dos Dados:

- **Renomeação de Colunas:** As colunas dos conjuntos de dados foram renomeadas para usar nomes mais descritivos em português, facilitando a compreensão.
- **Tratamento de Valores Ausentes:** Valores ausentes foram tratados de forma específica para cada tipo de dado. Colunas numéricas tiveram seus valores ausentes preenchidos com a média da coluna, enquanto colunas categóricas foram preenchidas com a categoria mais frequente.
- **Remoção de Duplicatas:** Registros duplicados foram identificados e removidos para evitar distorções na análise.
- **Exclusão de Colunas Irrelevantes:** Colunas consideradas irrelevantes para a análise, como IDs de programas e datas de promoções, foram excluídas.

3. Integração dos Dados:

- Os três conjuntos de dados foram mesclados em um único conjunto de dados, utilizando o ID do empregado como chave para conectar as informações. Essa integração permitiu uma visão mais completa da experiência dos funcionários.

4. Armazenamento dos Dados:

- O conjunto de dados final, após a limpeza e integração, foi salvo em um novo arquivo CSV para uso em análises futuras. Um diretório dedicado foi criado para armazenar os dados processados.

Resultados:

O processo de análise resultou em um conjunto de dados limpo, integrado e organizado, pronto para ser utilizado em análises mais aprofundadas sobre a experiência dos funcionários.

Interpretação dos Resultados:

- A limpeza e organização dos dados garantem a qualidade e confiabilidade das informações, permitindo insights mais precisos.
- A integração dos dados fornece uma visão holística da experiência dos funcionários, considerando diferentes aspectos como saúde mental, bem-estar e progressão de carreira.
- O conjunto de dados final é uma base sólida para futuras análises, permitindo a investigação de relações entre diferentes variáveis e a identificação de tendências importantes.

Próximos Passos:

Com o conjunto de dados preparado, é possível realizar análises mais aprofundadas para explorar questões como:

- Impacto do trabalho remoto na saúde mental e bem-estar dos funcionários.
- Eficácia dos programas de bem-estar oferecidos pela empresa.
- Fatores que influenciam a progressão de carreira dos funcionários.

Essas análises podem fornecer insights valiosos para a tomada de decisões estratégicas, visando melhorar a experiência dos funcionários e promover um ambiente de trabalho mais saudável e produtivo.

Relatório da Análise de Dados de Saúde Mental e Trabalho Remoto

1. Introdução

Este relatório apresenta uma análise exploratória de dados relacionados à saúde mental e ao trabalho remoto. O objetivo é identificar padrões e insights sobre o impacto do trabalho remoto na saúde mental dos colaboradores, considerando variáveis como nível de estresse, condições de saúde mental, satisfação com o trabalho remoto, suporte da empresa e outros fatores relevantes.

2. Metodologia

A análise foi realizada utilizando a linguagem de programação Python e bibliotecas como Pandas, NumPy e Plotly. Os dados foram carregados a partir de um arquivo CSV e submetidos a um processo de análise exploratória, que incluiu:

- **Análise Descritiva:** Cálculo de medidas estatísticas descritivas, como média, desvio padrão, quartis, etc., para cada variável numérica.
- **Visualização de Dados:** Criação de gráficos, como histogramas e gráficos de barras, para visualizar a distribuição dos dados e identificar padrões.
- **Análise de Relações:** Investigação da relação entre diferentes variáveis, como gênero, cargo, indústria e saúde mental, utilizando gráficos de barras agrupadas e boxplots.

3. Resultados e Discussão

3.1. Análise Descritiva

- **Idade:** A média de idade dos colaboradores é de 41 anos, com uma variação considerável entre 22 e 60 anos.

- **Experiência:** A média de anos de experiência é de 17,81 anos, indicando uma força de trabalho experiente.
- **Horas Trabalhadas:** A média de horas trabalhadas por semana é de 39,61 horas, próxima à jornada de trabalho típica de 40 horas.
- **Reuniões Virtuais:** A média de reuniões virtuais semanais é de 7,56, com variação de 0 a 15 reuniões.
- **Equilíbrio Vida-Trabalho:** A avaliação média do equilíbrio vida-trabalho é de 2,98 (em uma escala de 1 a 5), indicando uma percepção moderada ou insatisfeita.
- **Isolamento Social:** A avaliação média de isolamento social é de 2,99, similar à do equilíbrio vida-trabalho.
- **Suporte da Empresa:** A avaliação média do suporte da empresa no trabalho remoto é de 3,01, sugerindo uma percepção neutra ou levemente positiva.

Interpretação: Os dados indicam que a maioria dos colaboradores trabalha em torno de 40 horas semanais, com uma quantidade significativa de reuniões virtuais. As avaliações de equilíbrio vida-trabalho e isolamento social são moderadas, sugerindo que há espaço para melhorias no ambiente de trabalho remoto. O suporte da empresa é percebido de forma neutra ou levemente positiva.

3.2. Análise de Relações

- **Relação entre Gênero e Saúde Mental:** A análise revelou que a distribuição de gênero em relação ao nível de estresse, condição de saúde mental, satisfação com o trabalho remoto, atividade física e qualidade do sono é relativamente equilibrada. Não foram observadas diferenças significativas entre os gêneros.
- **Relação entre Cargo e Saúde Mental:** A análise indicou que o cargo pode influenciar a saúde mental dos colaboradores. Por exemplo, alguns cargos apresentam maior nível de estresse e menor satisfação com o trabalho remoto.
- **Relação entre Indústria e Saúde Mental:** A análise mostrou que a indústria em que o colaborador trabalha pode ter um impacto na sua saúde mental. Algumas indústrias apresentam maior nível de estresse e menor satisfação com o trabalho remoto.

- **Relação entre Local de Trabalho e Saúde Mental:** A análise revelou que o local de trabalho pode influenciar a saúde mental dos colaboradores. Por exemplo, colaboradores que trabalham em casa podem apresentar maior nível de isolamento social.
- **Relação entre Cargo e Horas Trabalhadas:** A análise indicou que a média de horas trabalhadas por semana é semelhante entre os diferentes cargos, em torno de 40 horas.

Interpretação: A análise de relações indica que fatores como cargo, indústria e local de trabalho podem influenciar a saúde mental dos colaboradores. É importante que as empresas considerem esses fatores ao implementar políticas de trabalho remoto e oferecer suporte aos seus colaboradores.

4. Conclusões e Recomendações

- O trabalho remoto pode ter um impacto significativo na saúde mental dos colaboradores, tanto positivo quanto negativo.
- É importante que as empresas ofereçam suporte aos seus colaboradores para garantir um ambiente de trabalho remoto saudável e produtivo.
- As empresas devem considerar fatores como cargo, indústria e local de trabalho ao implementar políticas de trabalho remoto.
- É fundamental promover o equilíbrio vida-trabalho, reduzir o isolamento social e oferecer acesso a recursos de saúde mental para os colaboradores em trabalho remoto.
- A análise exploratória de dados pode ser uma ferramenta valiosa para identificar padrões e insights sobre a saúde mental e o trabalho remoto.

Relatório sobre a Modelagem Preditiva

O código realiza análises preditivas e de clusterização usando dados de funcionários. Ele aborda várias questões de pesquisa relacionadas ao bem-estar, saúde mental e satisfação no trabalho, especialmente no contexto do trabalho remoto.

1. Previsão do Nível de Estresse:

- **Objetivo:** Prever o nível de estresse auto-relatado com base em características demográficas, local de trabalho e feedback de programas de bem-estar.
- **Método:** Usa o algoritmo RandomForestClassifier para construir um modelo preditivo.
- **Variáveis:** Região, Local de Trabalho, Pontuação de Feedback (independentes); Nível de Estresse (dependente).
- **Resultados:** O modelo teve um desempenho relativamente baixo, o que indica que essas variáveis, por si só, podem não ser suficientes para prever o nível de estresse com precisão.
- **Conclusões:** A baixa precisão do modelo ressalta a complexidade do estresse e a necessidade de considerar outros fatores. A baixa correlação entre o nível de estresse com a Região, Pontuação do FeedBack e Local de Trabalho confirmam essa conclusão.

2. Correlação entre Participação em Programas de Bem-Estar e Satisfação com o Trabalho Remoto:

- **Objetivo:** Investigar a correlação entre a participação em programas de bem-estar e a satisfação com o trabalho remoto.
- **Método:** Calcula a correlação de Spearman e usa um modelo de regressão linear para prever a satisfação.
- **Variáveis:** Status de Participação, Acesso a Recursos de Saúde Mental, Atividade Física, Qualidade do Sono (independentes); Satisfação com o Trabalho Remoto (dependente).
- **Resultados:** A correlação entre a participação em programas de bem-estar e a satisfação no trabalho remoto foi baixa. O modelo

preditivo identificou a participação como um fator importante para a satisfação.

- Conclusões: Programas de bem-estar podem ter um impacto positivo na satisfação com o trabalho remoto, mas outros fatores provavelmente também estão envolvidos.

3. Previsão de Condições de Saúde Mental:

- Objetivo: Prever a probabilidade de um empregado relatar ansiedade ou depressão com base em idade, gênero, carga de trabalho e interação social.
- Método: Usa o algoritmo RandomForestClassifier para construir um modelo preditivo.
- Variáveis: Idade, Gênero, Horas Trabalhadas por Semana, Avaliação de Isolamento Social (independentes); Condição de Saúde Mental (dependente).
- Resultados: O modelo obteve uma precisão razoável, sendo capaz de identificar funcionários em risco de condições de saúde mental.
- Conclusões: Variáveis como idade, gênero, carga de trabalho e interação social podem ser usadas para prever condições de saúde mental.

4. Impacto das Mudanças na Carreira na Saúde Mental:

- Objetivo: Analisar se as mudanças na carreira impactam a saúde mental.
- Método: Usa o algoritmo RandomForestClassifier para construir um modelo preditivo.
- Variáveis: Melhoria Observada, Cargo Anterior, Novo Cargo, Mudança Salarial, Satisfação com o Cargo (independentes); Condição de Saúde Mental (dependente).
- Resultados: O modelo obteve uma precisão moderada na previsão de mudanças na saúde mental.
- Conclusões: Mudanças na carreira podem influenciar a saúde mental dos funcionários, mas mais pesquisas são necessárias para confirmar a relação.

5. Fatores que Contribuem para o Isolamento Social:

- **Objetivo:** Identificar os fatores que mais contribuem para a sensação de isolamento social entre funcionários remotos.
- **Método:** Usa o algoritmo RandomForestClassifier para construir um modelo preditivo.
- **Variáveis:** Gênero, Nível de Estresse, Condição de Saúde Mental, Satisfação com o Trabalho Remoto, Atividade Física, Qualidade do Sono, Status de Participação, Suporte da Empresa ao Trabalho Remoto (independentes); Avaliação de Isolamento Social (dependente).
- **Resultados:** O modelo obteve uma precisão razoável, identificando funcionários em risco de alta isolação social.
- **Conclusões:** Nível de estresse, satisfação com o trabalho remoto, suporte da empresa e outros fatores podem contribuir para o isolamento social.

6. Eficácia dos Programas de Bem-Estar:

- **Objetivo:** Avaliar a eficácia dos programas de bem-estar com base no feedback dos funcionários e scores de melhoria.
- **Método:** Usa o algoritmo RandomForestClassifier para construir um modelo preditivo.
- **Variáveis:** Pontuação de Feedback, Melhoria Observada (independentes); Nome do Programa (dependente).
- **Resultados:** O modelo conseguiu prever a eficácia de futuros programas, mas com ressalvas sobre o programa "Work-Life Balance Workshop".
- **Conclusões:** O modelo pode ser útil para orientar a escolha de programas de bem-estar.

Análise de Clusterização com K-means

Clusterização de Empregados:

- Objetivo: Agrupar funcionários com base em suas características demográficas, níveis de estresse e satisfação com o trabalho remoto.
- Método: Usa o algoritmo K-means para clusterizar os dados.
- Variáveis: Idade, Gênero, Região, Nível de Estresse, Mudança Salarial, Satisfação com o Trabalho Remoto.
- Resultados: O código clusterizou os funcionários em 3 grupos distintos, visualizados em um gráfico de dispersão.
- Conclusões: Os clusters revelam padrões e diferenças entre os grupos de funcionários, fornecendo insights sobre como as condições de trabalho afetam diferentes grupos.

Observações Gerais:

- O código está bem estruturado e documentado.
- As análises são relevantes para entender o bem-estar e a saúde mental dos funcionários.
- Os modelos preditivos têm precisão variável, mas fornecem insights importantes.
- A análise de clusterização oferece uma visão geral da diversidade dos funcionários.
- É importante lembrar que os resultados são baseados em um conjunto de dados específico e podem não ser generalizáveis para outras populações.

Links

[Github](#)

[BigQuery](#)

[Dashboard](#)