Introdução e Objetivo:

O dataset sintético trabalhado consiste em dados sobre a produtividade e bem-estar no trabalho remoto e presencial. Formando um dataset de 1000 funcionários (1000 linhas) contendo informações como id, tipo de emprego, horas trabalhadas na semana, score de produtividade e score de bem-estar, no caso as colunas.

O objetivo do projeto e analisar o impacto dos tipos de trabalhos na produtividade e no bem-estar, identificar as diferenças e analisar correlação entre colunas, e além de analisar o impacto, também comprovar de diversas formas que resultado desses impactos existem.

Organização dos arquivos:

ArquivoPowerBI 🡪 Local onde está a representação gráfica do projeto.

Documentation 🡪 Local onde está a documentação do projeto e explicação.

Main 🡪 Local onde está o script da análise, no arquivo “main.ipynb”.

\--> database 🡪 Local onde está a base de dados utilizada neste projeto

**Coleta de Dados**:

A origem dos dados vem do site da kaggle: <https://www.kaggle.com/> . Site esse que hospeda uma vasta coleção de conjuntos de dados públicos que podem ser usados para projetos de aprendizado de máquina e análise de dados. Esse dataset em específico foi disponibilizado como arquivo em formato csv.

Exploração e Limpeza de Dados:

As únicas limpeza nos dados foram:

- Nome da variáveis da coluna “employment\_type” minúsculo.

- Colando “employee\_id” como index

- Nome das colunas minúsculos.

- Tirando linhas com valores “Nan”.

**Transformação de Dados**:

As principais técnicas de manipulação usadas foram a separação por tipo de emprego criando 2 df’s remote e in-office, para uma análise individual de cada cenário. Ao decorrer do arquivo “main.ipynb” contém algumas outras separações e agregações básicas nos dados.

Com o auxilio das bibliotecas, pandas e NumPy.

**Visualização de Dados:**

Ao decorrer da análise foram criados gráficos para comprovar teste estatísticos, e também confirmar números e médias. Como por exemplo um BoxPlot para mostrar e aumentar a eficiência do ttest feito.

E foi criado também a parte a parte gráfica do projeto pelo PowerBI aonde é mostrado de forma visual algumas análise estatísticas, por meio de gráfico de barra, dispersão, pizza e gráfico de área.

**Conclusão:**

As conclusões tiradas a partir da análise foram com base na comparação de produtividade e bem-estar dos funcionários partindo da separação entre remoto e no-escritório (presencial).

As conclusões tiradas foram:

Um funcionário médio no estilo remoto tem 15.45% a mais de produtividade e 18.44% a mais de bem-estar que o funcionário no estilo presencial. Além do funcionário remoto ter essas médias maiores ele trabalha cerca de 27.59% a menos em relação a horas semanais do que o funcionário presencial.

Tiramos do ttest que há uma diferença significativa entre as media remoto produtividade e presencial e produtividade.

Foi provado que não há conrrelação entre horas trabalhadas e produtividade, em ambos os cenários, tanto presencial quanto remoto. E também que não há correlação quando falamos de produtividade x bem-estar.

Em resumo, voltando ao titulo do projeto “Trabalho Remoto e Presencial: Calculando suas diferenças e mostrando o proveito do melhor.” Foi concluído no projeto um melhor desenvolvimento de produtividade e bem-estar no trabalho remoto, expressando seus proveitos e suas médias.

Referências:

Sites:

<https://www.kaggle.com/>

<https://pandas.pydata.org/docs/reference/api/pandas.melt.html>

<https://seaborn.pydata.org/generated/seaborn.boxplot.html>

Tecnologias utilizadas e Bibliotecas:

Pandas, Matplotlib, SciPy, Seaborn, PowerBI