



Documentação - Pipeline - Projeto Final

- **Objetivo:**
 - **Explorar indicadores** de rotatividade (attrition) de funcionários e **construir um dashboard** interativo que ajudem a empresa fictícia **Data Girls S.A** a tomar decisões estratégicas de RH:
- **Perguntas Norteadoras** de Negócio:
 - Qual perfil tem maior risco de rotatividade?
 - Quais fatores (salário, idade, overtime etc.) mais influenciam na decisão de sair?
 - Qual departamento/cargo merece atenção imediata?
 - Há padrão na duração do tempo de casa antes da saída?
- **Etapas** orientadoras do Projeto:
 1. Leitura e Exploração Inicial
 2. Limpeza e Preparação dos Dados
 3. Análise Exploratória (EDA)
 4. Visualização Interativa / Dashboard
 5. Insights e Recomendações
 - Identifique os principais ofensores.
 - Sugira políticas, por exemplo:
 - Controle de horas extras,
 - Incentivo à proximidade ao trabalho,
 - Programas de satisfação ou treinamentos.

1. Leitura e Exploração Inicial:

- a. A linguagem escolhida foi Python, mesmo pelos conhecimentos básicos, há uma maior familiaridade;
- b. A análise será realizada no Colab (Notebook Jupyter) - [Link](#);
- c. Organização das **variáveis** do dataset disponibilizado - [Link](#);
- d. Dataset possui **1470 linhas e 35 colunas**
- e. Importação de biblioteca e leitura de dados
 - i. **Aula referência** - Introdução ao Pandas (Mirla Borges) - [Link](#)
 - ii. Feito a importação do dataset através do drive do Google;
 - iii. Leitura dos dados e identificação dos tipos;
 1. Na página do dataset ([Link](#)) há sinalizados os tipos:

35 columns	
Integer	26
String	6
Boolean	3

2. No python há:
 - a. dtypes: int64 (26), object (9)

2. Limpeza e Preparação dos Dados:

- a. Identificação de **Nulos** / Ausentes:
 - i. **Aula referência** - Introdução ao Pandas (Mirla Borges) - [Link](#)
 - ii. Não há nulos no dataset disponibilizado;
- b. Tratando o **tipo** das variáveis:
 - i. **Aula referência** - Limp. e transformação de dado (Laura D.) - [Link](#)
 - ii. Sem necessidade, todos os tipos em conformidade;
- c. Identificando dados duplicados:
 - i. **Referência** - adaptei a fórmula para achar nulos;
 - ii. Não há duplicados no dataset disponibilizado;
- d. Identificando outliers:
 - i. **Aula referência** - Notebook/Colab - (Laura D.) - [Link](#)
 - ii. Variáveis numéricas analisadas:
 - 1.

Age
DailyRate
DistanceFromHome

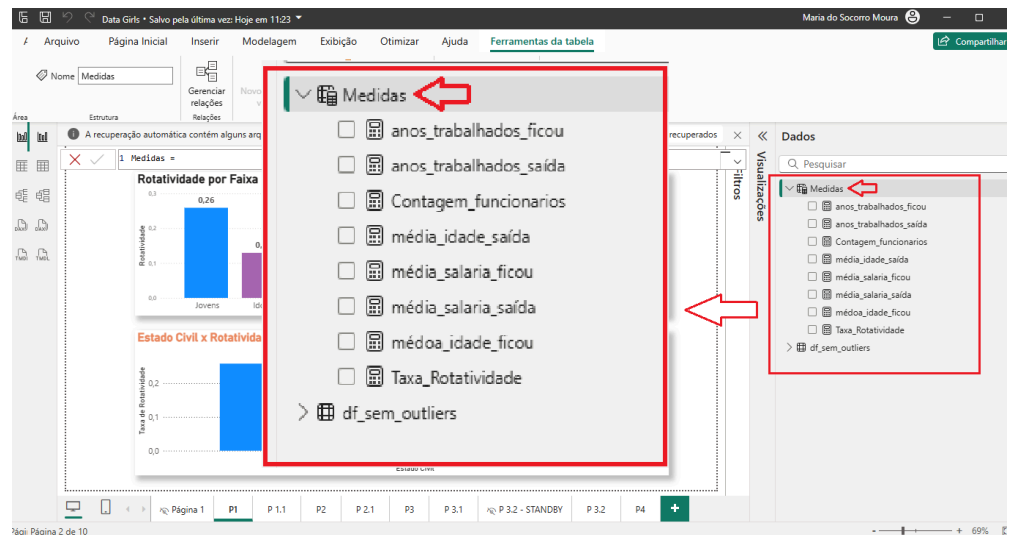
HourlyRate
MonthlyIncome
MonthlyRate
NumCompaniesWorked
PercentSalaryHike
TotalWorkingYears
TrainingTimesLastYear
YearsAtCompany
YearsInCurrentRole
YearsSinceLastPromotion
YearsWithCurrManager

- iii. As variáveis destacadas com laranja foram identificados outliers;
- iv. Optei pela remoção dos outliers , tendo como referência o material da instrutora e por entender que para o objetivo da análise, os outliers foram interpretados como funcionários mais estáveis em vários aspectos ("Total de Anos Trabalhados", "Número total de anos na empresa", "Anos na função atual"...), e por esse motivo podem distorcer as análises relacionadas à rotatividade.
- v. Organização da análise de **outliers** - [Link \(Doc\)](#)

3. Análise Exploratória (EDA):

- a. Para essa fase, vou baixar o arquivo tratado no Colab e importar para o Power BI, e assim iniciar a análise exploratória;
- b. **Aula referência** - Construção de Dashboards Profissionais (Leda G.) - [Link](#);
- c. **Aula referência** - Power BI - Dashboards Interativos (Ana Duarte) - [Link](#);
- d. **Aula referência** - Power BI - Introdução ao Power BI (Isabela Troyo) - [Link](#);
- i. Iniciei com a criação de medidas, as quais entendi que poderia fazer de forma mais organizada segundo orientação em vídeo:

ii.



iii. As fórmulas aplicadas e destacadas acima foram feitas com **auxílio de Inteligência Artificial (IA - Gemini)**:

1.

```
anos_trabalhados_ficou =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[TotalWorkingYears]),  
    'df_sem_outliers'[Attrition] = "No"  
)
```

2.

```
anos_trabalhados_saida =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[TotalWorkingYears]),  
    'df_sem_outliers'[Attrition] = "Yes"  
)
```

3.

```
Contagem_funcionarios =  
COUNT('df_sem_outliers'[EmployeeNumber])
```

4.

```
média_idade_saida =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[Age]),  
    'df_sem_outliers'[Attrition] = "Yes"  
)
```

5.

```
média_salario_ficou =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[MonthlyIncome]),  
    'df_sem_outliers'[Attrition] = "No"  
)
```

6.

```
média_salario_saida =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[MonthlyIncome]),  
    'df_sem_outliers'[Attrition] = "Yes"  
)
```

7.

```
médida_idade_ficou =  
CALCULATE(  
    AVERAGE('df_sem_outliers'[Age]),  
    'df_sem_outliers'[Attrition] = "No"  
)
```

8.

```
Taxa_Rotatividade =  
DIVIDE(  
    CALCULATE(COUNTROWS('df_sem_outliers'),  
        'df_sem_outliers'[Attrition] = "Yes"),  
    COUNTROWS('df_sem_outliers')  
)
```

iv. Assim, para produzir novos insights , agrupei dados de acordo com variáveis categóricas:

1. Assim, foi utilizada a seguinte fórmula:

2.

```
Faixa Etária =  
IF('df_sem_outliers'[Age] <= 30, "Jovens",  
    IF('df_sem_outliers'[Age] <= 50, "Adultos",  
        "Idosos"  
    )  
)
```

4. Visualização Interativa / Dashboard:

- a. **Aula referência** - Construção de Dashboards Profissionais (Leda G.) - [Link](#);
- b. **Aula referência** - Power BI - Dashboards Interativos (Ana Duarte) - [Link](#);
- c. **Aula referência** - Power BI - Introdução ao Power BI (Isabela Troyo) - [Link](#);
- i. Para essa fase, temos a produção de **dashboards** diversos que dão uma sequência lógica para o que se deseja contar através dos dados:
 - 1. Aqui está o arquivo em **PDF** exportado diretamente do **Power BI** - [Link](#);
 - 2. Aqui segue também a **apresentação** que insere o devido **contexto** aos dados coletados e analisados - [Link](#);
 - 3. O arquivo de apresentação pode ser utilizado como **guia de uso**;
 - 4. Para a contextualização foi utilizada a técnica de **storytelling**.

5. Insights e Recomendações

- a. **Insights** - Identificação dos principais **ofensores**. A IA - Gemini foi usada de forma a auxiliar nas complementações, embora o mecanismo não se mostrou muito sensível aos dados referentes às mulheres, por isso a necessidade de edição e revisão constantes. Segue principais descobertas:
 - i. **Fatores de Carga de Trabalho:** A prática de horas extras é o maior preditor de rotatividade, com uma taxa de saída três vezes maior para quem a pratica.
 - ii. **Problema de Início de Carreira:** A rotatividade é altíssima no primeiro ano de empresa e entre a faixa etária jovem e com menos anos de experiência.
 - iii. **Remuneração:** O salário mediano dos funcionários que saem é significativamente menor do que o dos que ficam, indicando um problema de competitividade salarial.
 - iv. **Departamentos/Cargos de Risco:** O departamento de Vendas e, especificamente, os cargos de Representante de Vendas e Técnico de Laboratório apresentam taxas de rotatividade alarmantes.
 - v. **Diferença de Gênero:** A taxa de rotatividade para o gênero masculino é ligeiramente superior, mas a taxa feminina também é

alta, indicando que o problema afeta a todos. Basta observar, como destacado anteriormente, que no cargo de Representante de Vendas e Recursos Humanos, a taxa de saída para as mulheres é maior do que para os homens! **Por que isso acontece?**

b. Recomendações:

i. Para **todos e todas**:

1. **Política:** Implementar um teto de horas extras e monitorar ativamente os cargos de alto risco (Sales Representative, Laboratory Technician) para evitar esgotamento:
 - a. **Ação:** Lançar uma pesquisa de clima focada na carga de trabalho e na percepção de equilíbrio.;
2. **Política:** Focar os esforços de retenção nos primeiros 12 a 18 meses de contratação:
 - a. **Ação:** Implementar um programa de mentoria obrigatório para novos funcionários, pareando-os com colegas mais experientes para suporte e integração. Oferecer planos de desenvolvimento de carreira claros e personalizados para os cargos de alto risco, mostrando um caminho de crescimento;

ii. Específicas para o **Público Feminino**:

1. Apesar de a taxa de rotatividade masculina ser ligeiramente superior, o problema também é **relevante** para as **mulheres**. Para a gestão, o foco deve ser na retenção de todo o talento. Sugestões específicas podem incluir:
 - a. **Política:** Criar uma "Força-Tarefa de Retenção Feminina" para entender as causas específicas de saída:
 - i. **Ação:** Realizar entrevistas de saída focadas para entender as razões da rotatividade feminina e instituir programas de mentoria exclusivos para mulheres, auxiliando-as no desenvolvimento de carreira e na superação de desafios específicos.
 - ii. **Ação:** Revisar as políticas de horários flexíveis e trabalho híbrido para melhor atender às necessidades das funcionárias, o que também beneficiará a todos.