

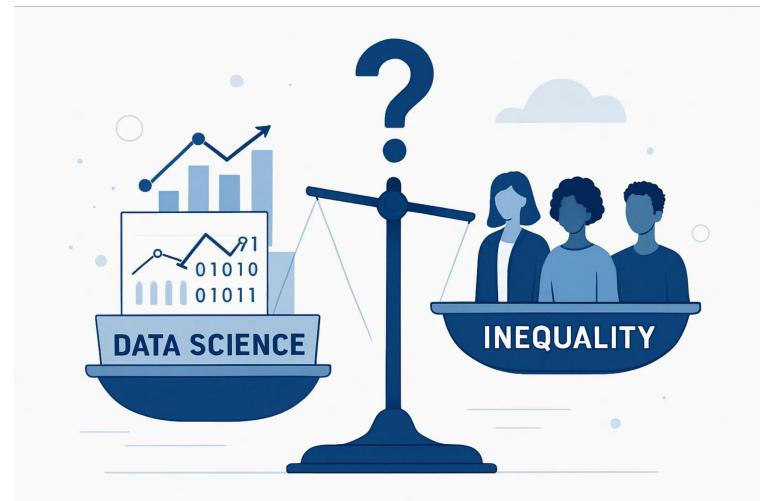
# Se a ciência de dados é sobre a realidade, por que ela pode ser cega para a desigualdade?

## Ciência de Dados, Desigualdade e Interseccionalidade

Os dados, mesmo sendo números, **não são neutros**. Eles refletem a sociedade que os produz, incluindo suas estruturas de desigualdade.

Nesta aula, vamos explorar como as desigualdades estruturais se manifestam nos dados e como a análise crítica pode nos ajudar a identificar e mitigar esses problemas.

*"Os dados não falam por si mesmos. Eles falam por quem os coleta, analisa e interpreta."*



### Aula 8

Ensino Médio/Técnico (intermediário)

# O que é Interseccionalidade?

## Definição

Interseccionalidade é a ideia de que diferentes identidades sociais (como raça, gênero, classe, sexualidade) se cruzam, criando experiências únicas de opressão e privilégio que não podem ser entendidas olhando para cada categoria isoladamente.

⚙️ **Múltiplas dimensões** - Uma pessoa não é apenas definida por sua raça, ou apenas por seu gênero, mas pela combinação de todas as suas identidades sociais.

📦 **Efeitos combinados** - As desvantagens não são simplesmente somadas, mas interagem de formas complexas e únicas.

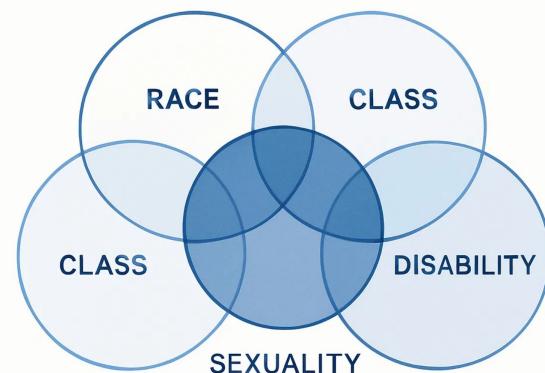
🔍 **Análise mais completa** - Permite uma compreensão mais profunda das desigualdades estruturais e como elas afetam diferentes grupos.

## Exemplo na Ciência de Dados:

Ao analisar disparidades salariais, não basta olhar separadamente para gênero ou raça. Uma mulher negra pode enfrentar desvantagens diferentes e mais intensas do que uma mulher branca ou um homem negro.

*Termo cunhado pela jurista Kimberlé Crenshaw em 1989 para explicar como mulheres negras enfrentavam discriminação tanto por raça quanto por gênero.*

## INTERSECTIONALITY



# Desigualdades Estruturais

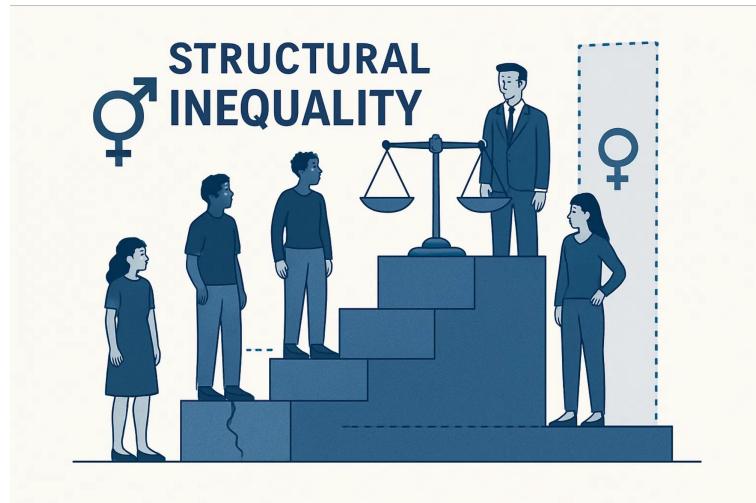
## O que são Desigualdades Estruturais?

- ⌚ **Construção histórica** - Desigualdades de raça, gênero e classe não são acidentais. São construídas historicamente através de políticas, práticas e normas sociais.
- ▣ **Reforço institucional** - São reforçadas por sistemas sociais e institucionais, como educação, mercado de trabalho, sistema de justiça e saúde.
- ↳ **Reprodução nos dados** - Os dados, coletados dentro desses sistemas, muitas vezes reproduzem e até amplificam essas desigualdades.

## Por que isso importa na Ciência de Dados?

- ⌚ Algoritmos treinados com dados enviesados tendem a perpetuar e até amplificar as desigualdades existentes.
- ⌚ Sistemas automatizados de decisão podem parecer "objetivos", mas podem reforçar padrões discriminatórios.

"Não podemos resolver problemas usando o mesmo tipo de pensamento que usamos quando os criamos." - Albert Einstein



# Viés de Amostra

## O que é Viés de Amostra?

Ocorre quando os dados coletados não representam de forma justa a população que o algoritmo deve servir, levando a conclusões erradas e sistemas que funcionam melhor para alguns grupos do que para outros.

## ❤️ Exemplo: Sistemas de Saúde

Um sistema de saúde baseado em dados majoritariamente de um grupo étnico pode ser ineficaz para outros grupos, pois:

- ⚠ Diferentes grupos étnicos podem apresentar **diferentes fatores de risco** para certas doenças.
- ⚠ Sintomas podem se **manifestar de formas diferentes** dependendo da etnia, gênero ou idade.
- ⚠ A **eficácia de tratamentos** pode variar entre diferentes grupos populacionais.

## Como mitigar o viés de amostra?

Garantir diversidade nos dados de treinamento

Testar algoritmos com dados representativos de diferentes grupos

Incluir especialistas diversos no desenvolvimento e validação

Monitorar continuamente o desempenho do sistema para diferentes grupos



# Desigualdade Salarial

## Como os dados revelam desigualdades salariais

 **Mesma função, salários diferentes** - Os dados podem revelar diferenças salariais entre gêneros ou raças, mesmo quando analisamos pessoas com as mesmas funções, qualificações e experiência.

 **Efeito interseccional** - Análise de dados mostra que mulheres negras frequentemente enfrentam as maiores disparidades salariais, evidenciando o efeito combinado de gênero e raça.

 **Além da média** - Análises mais sofisticadas podem revelar como as disparidades aumentam em posições de maior hierarquia ou em setores específicos da economia.

### Exemplo de análise:

Ao analisar dados salariais de uma empresa, podemos agrupar por gênero e raça usando:

```
df.groupby(['genero', 'raça'])['salario'].mean()
```

Este tipo de análise pode revelar disparidades que ficariam ocultas em análises mais simples.

*Nota: Estudos do IBGE mostram que mulheres ganham, em média, cerca de 20% menos que homens no Brasil, mesmo em funções equivalentes.*



# Representatividade

## Como os dados revelam representatividade?

**⚖️ Comparação demográfica** - Os dados podem ser usados para comparar a composição de uma organização (por gênero, raça, idade) com a composição da população geral.

**📊 Distribuição por níveis** - Além da representação geral, é importante analisar a distribuição em diferentes níveis hierárquicos (ex: liderança vs. entrada).

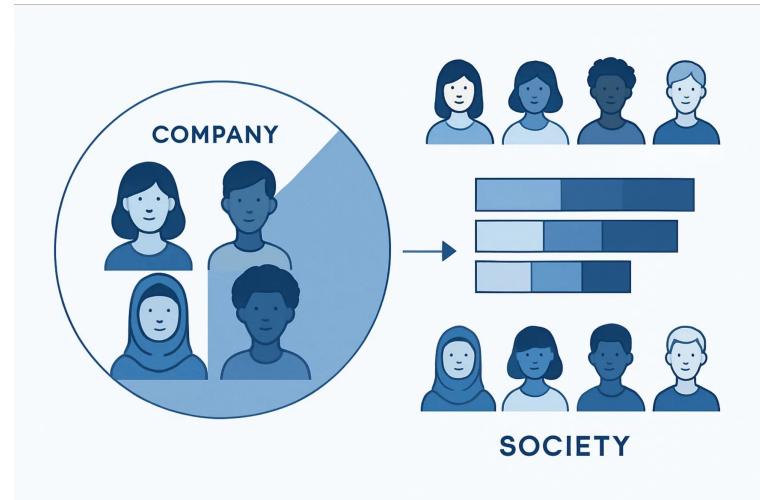
**⬇️ Tendências ao longo do tempo** - A análise de dados históricos permite identificar se a representatividade está melhorando, piorando ou estagnada.

### Exemplo prático:

Uma empresa com 30% de mulheres em seu quadro geral, mas apenas 5% em cargos de liderança, revela um problema de representatividade nos níveis hierárquicos superiores, mesmo que pareça diversa no total.

### 💡 Para reflexão:

Quais outros exemplos de desigualdade você vê no dia a dia que poderiam ser revelados através da análise de dados?



# Atividade A: Análise de Dados em Grupos

## Análise dos Dados

Divida a turma em grupos. Forneça a cada grupo um dataset simulado de RH com variáveis de gênero, raça, salário e cargo.

### 1. Representatividade

Calcule a proporção de cada grupo de gênero e raça no dataset.

```
# Contagem por gênero  
df['genero'].value_counts(normalize=True)
```

### 2. Análise Salarial

Calcule a média salarial para cada grupo.

```
# Média salarial por gênero  
df.groupby('genero')['salario'].mean()
```

### 3. Análise Interseccional

Combine as variáveis para análise interseccional.

```
# Média salarial por gênero e raça  
df.groupby(['genero', 'raça'])['salario'].mean()
```



#### Dica

Vale destacar como a análise interseccional revela desigualdades ocultas nas análises separadas.

# Atividade B: Discussão sobre Vieses e Desigualdades

## Debate em Grupo

### Instruções:

Oriente um debate em grupo com base na análise de dados realizada na Atividade A. Reflita criticamente sobre os resultados encontrados.

### Perguntas para o Debate:

- ? "Esses números refletem um problema de hoje ou um problema histórico?"
- ? "O que a análise de dados por si só não nos conta sobre as causas dessas disparidades?"
- ? "Qual é o limite dessa análise? O que precisamos saber além dos dados para entender a realidade?"



### Dicas

- Vale pensar criticamente sobre os dados
- Destacar a importância do contexto histórico e social
- Promover um ambiente seguro para que todos possam expressar suas opiniões
- Conectar com exemplos do cotidiano

# Atividade C: Construção de Relatórios Diagnósticos

## Construção de Relatórios Diagnósticos

### Instruções:

Com base na análise da Atividade A, cada grupo deve construir um pequeno relatório diagnóstico em Google Docs ou Slides.

### Roteiro do Relatório:

**H** **Título:** "Relatório de Disparidades na Empresa X"

**L** **Gráficos de Representatividade:** Incluir gráficos de barras para mostrar a proporção de cada grupo de gênero e raça.

**L** **Análise de Disparidades:** Incluir um gráfico de barras comparando a média salarial entre os grupos (gênero, raça e interseccional).

**D** **Texto Explicativo:** Escrever um texto simples explicando as principais descobertas.



### Exemplo de texto para o relatório:

"A análise mostra que, embora as mulheres representem X% da equipe, seu salário médio é Y% menor do que o dos homens. Quando analisamos por raça, observamos que pessoas negras recebem em média Z% menos que pessoas brancas na mesma função."

```
# Código para gerar o gráfico de barras
import matplotlib.pyplot as plt
salarios = df.groupby(['genero', 'raça'])['salario'].mean()
salarios.plot(kind='bar')
plt.title('Média Salarial por Gênero e Raça')
plt.ylabel('Salário Médio (R$)')
plt.savefig('disparidade_salarial.png')
```

# Atividade D: Propostas de Ação

## Propostas de Ação para Equidade

### Instruções:

Com base no relatório diagnóstico criado na atividade anterior, os grupos devem sugerir pelo menos duas ações que a "empresa" poderia tomar para promover a equidade.

- 1 Identifique os principais problemas revelados pelos dados
- 2 Proponha ações concretas e viáveis para resolver esses problemas
- 3 Considere o impacto potencial e a viabilidade de implementação

### Exemplos de Ações:

- ⟳ "Revisar a política de salários para garantir equidade entre gêneros e raças para funções equivalentes."
- 人群 "Criar um programa de mentoria para grupos sub-representados em cargos de liderança."
- brain "Implementar treinamento de vieses inconscientes para gestores responsáveis por contratações e promoções."
- checklist "Estabelecer metas de diversidade com prazos definidos e monitoramento regular."



### Dica

Pensar em soluções que abordem as causas estruturais das desigualdades, não apenas seus sintomas. Destaque que a ciência de dados pode ser uma ferramenta poderosa para monitorar o progresso dessas ações ao longo do tempo.

# Atividade E: Apresentação e Troca de Feedback

## Apresentação dos Relatórios

### Instruções:

Cada grupo apresenta seu relatório diagnóstico e suas propostas de ação para a turma.

- 1 Cada grupo terá **3 minutos** para apresentar:
  - As principais disparidades encontradas nos dados
  - As propostas de ação sugeridas para promover equidade
- 2 Após cada apresentação, reserve **2 minutos** para feedback construtivo da turma.

### Perguntas para Feedback:

- » "Essa proposta de ação é viável? Por quê?"
- » "O que mais a análise poderia ter mostrado?"
- » "Como essa análise poderia ser aplicada em outros contextos?"



### Dicas

- Feedback construtivo e respeitoso
- Destaque pontos fortes de cada apresentação
- Promova conexões entre as diferentes análises
- Registre as principais ideias para a discussão final

# Encerramento e Reflexão

## O Papel e a Responsabilidade Social do Cientista de Dados

**Q Olhar crítico** - Como cientistas de dados, precisamos questionar constantemente: "Como nossa perspectiva sobre o mundo pode influenciar a forma como analisamos os dados?"

**A Responsabilidade ética** - Temos o dever de identificar e mitigar vieses em nossos dados e algoritmos, reconhecendo seu potencial impacto na perpetuação de desigualdades.

**IP Inclusão e diversidade** - Equipes diversas tendem a identificar problemas de viés mais facilmente e criar soluções mais abrangentes e justas.

*"A ciência de dados pode ser uma ferramenta poderosa para revelar desigualdades e promover a justiça social, mas apenas se usada com consciência crítica e responsabilidade."*



### Principais aprendizados da aula:

- ✓ Os dados podem revelar desigualdades estruturais
- ✓ A análise crítica e a contextualização são fundamentais
- ✓ A interseccionalidade nos ajuda a compreender experiências únicas de opressão
- ✓ A ciência de dados é uma ferramenta poderosa para promover a justiça social