

EDA Análise Exploratória dos Dados

1 Análise exploratória dos dados (EDA)

Neste estudo, o foco da EDA é examinar como as variáveis do conjunto de dados se relacionam com a variável alvo: **Diabetes**.

Para isso, utilizamos visualizações gráficas e cálculos estatísticos, como correlação e distribuição de frequências, a fim de identificar possíveis fatores associados ao diagnóstico de diabetes. Analisamos variáveis como pressão arterial, colesterol, atividade física, hábitos alimentares e condições de saúde pré-existentes para entender suas influências na presença da doença.

A partir dessa exploração inicial, podemos obter insights importantes que auxiliarão na formulação de recomendações baseadas nos dados.

2 Dicionário de variáveis:

Diabetes: 0 = sem diabetes, 1 = com diabetes

Pressao_Alta: 0 = sem pressão alta, 1 = com pressão alta

Colesterol_Alto: 0 = sem colesterol alto, 1 = com colesterol alto

Colesterol_Exame: 0 = não fez exame de colesterol na vida, 1 = fez exame de colesterol alguma vez

IMC: Índice de Massa Corporal (IMC)

Fumante: 0 = não fumante, 1 = fumante

AVC: 0 = sem histórico de AVC, 1 = com histórico de AVC

Problema_Cardiaco: 0 = sem histórico de doença cardíaca ou ataque cardíaco, 1 = com histórico de doença cardíaca ou ataque cardíaco

Atividade_Fisica: 0 = não pratica atividade física, 1 = pratica atividade física

Come_Frutas: 0 = não consome frutas, 1 = consome frutas

Come_Legumes: 0 = não consome vegetais, 1 = consome vegetais

Consumo_Bebida_Alcoolica: 0 = não consome álcool em altas quantidades, 1 = consome álcool em altas quantidades

Plano_Saude: 0 = não tem plano de saúde, 1 = tem plano de saúde

Sem_Dinheiro_Consultas: 0 = não foi ao médico por questões financeiras, 1 = foi ao médico por questões financeiras (últimos 12 meses)

Saude_Geral: Saúde geral (1 a 5) - 1 = Excelente, 2 = Muito boa, 3 = Boa, 4 = Aceitável, 5 = Ruim

Dias_Problemas_Mentais: Nos últimos 30 dias, quantos dias a saúde mental não foi boa (0 a 30)

Dias_Problemas_Fisicos: Nos últimos 30 dias, quantos dias a saúde física não foi boa (0 a 30)

Dificuldade_Andar: 0 = não tem dificuldade para caminhar, 1 = tem dificuldade para caminhar

Genero: 0 = feminino, 1 = masculino

Faixa_Idade: Idade em faixas 1 = 18-24; 2 = 25-29; 3 = 30-34; 4 = 35-39; 5 = 40-44; 6 = 45-49; 7 = 50-54; 8 = 55-59; 9 = 60-64; 10 = 65-69; 11 = 70-74; 12 = 75-79; 13 = 80+

Ensino: Níveis de ensino 1 = nunca frequentou a escola; 2 = escola primária; 3 = escola secundária incompleta ; 4 = escola secundária; 5 = faculdade incompleta ou curso técnico; 6 = completou faculdade ou diplomas superiores

Faixa_Renda: Renda anual em faixas 1 = < 10.000; 2 = 10.000-14.999; 3 = 15.000-19.999; 4 = 20.000-24.999; 5 = 25.000-34.999; 6 = 35.000-49.999; 7 = 50.000-74.999; 8 = 75.000+

3 Insights

3.1 Fatores de Saúde e Estilo de Vida

3.1.1 Pressão Alta vs. Diabetes

- 75.27% das pessoas com diabetes têm pressão alta, contra 37.42% dos não diabéticos.
- **Alto risco:** Pressão alta é uma das condições mais fortemente associadas ao diabetes.

3.1.2 Colesterol Alto vs. Diabetes

- 67.01% das pessoas com diabetes têm colesterol alto, em comparação com 38.13% dos não diabéticos.
- **Alerta:** Diabetes está relacionado a disfunções metabólicas que também afetam os níveis de colesterol.

3.1.3 Exame de Colesterol vs. Diabetes

- 99.32% das pessoas com diabetes fazem exames de colesterol, contra 95.73% dos não diabéticos.
- **Indica acompanhamento médico mais frequente**, essencial para controlar os riscos cardiovasculares do diabetes.

3.1.4 AVC vs. Diabetes

- 9.25% das pessoas com diabetes tiveram AVC, enquanto apenas 3.19% dos não diabéticos passaram por isso.
- **Diabetes aumenta o risco de AVC**, possivelmente por conta da hipertensão e problemas circulatórios.

3.1.5 Problema Cardíaco vs. Diabetes

- 22.29% das pessoas com diabetes têm doenças cardíacas, contra apenas 7.27% dos não diabéticos.
- **Confirma a forte relação entre diabetes e problemas cardíacos**

3.1.6 Atividade Física vs. Diabetes

- 63.05% das pessoas com diabetes praticam atividades físicas, contra 77.55% dos não diabéticos.
- A falta de atividade física está correlacionada ao diabetes

3.1.7 Alimentação (Frutas e Legumes) vs. Diabetes

- Menos diabéticos consomem frutas (58.54%) e legumes (75.64%) do que os não diabéticos (63.81% e 82.11%).

- Sugere que alimentação saudável pode ser um fator de proteção contra diabetes.

3.2 Condições Socioeconômicas e Saúde.

3.2.1 Plano de Saúde vs. Diabetes

- **95.98% das pessoas com diabetes possuem plano de saúde**, semelhante aos não diabéticos (95.01%).
- **Acesso ao plano de saúde não parece influenciar significativamente o risco de diabetes**, mas pode impactar o acompanhamento da doença.

3.3 Qualidade de Vida e Bem-Estar

3.3.1 Autoavaliação da Saúde vs. Diabetes

- Diabéticos se avaliam com saúde pior do que não diabéticos:
 - **"Ruim"**: 12.95% dos diabéticos vs. 3.48% dos não diabéticos.
 - **"Excelente"**: Apenas 3.23% dos diabéticos vs. 20.21% dos não diabéticos.
- **Sugere que o diabetes afeta negativamente a percepção da saúde pessoal.**

3.3.2 Dificuldade de Locomoção vs. Diabetes

- 37.12% das pessoas com diabetes têm dificuldades para andar, contra 13.42% dos não diabéticos.
- Indica impacto do diabetes na mobilidade, pessoas sedentárias são mais propensas a serem diabéticas

3.4 Demografia: Idade, Gênero, Educação e Renda

3.4.1 Idade vs. Diabetes

- O diabetes se torna mais comum com a idade:
- A maior incidência ocorre entre 60-79 anos, com pico na faixa 65-69 anos (18.55%).
- Reflete o fator etário como um grande risco para o diabetes

3.4.2 Gênero vs. Diabetes

- Percentual de diabetes é levemente maior entre homens (47.91%) do que entre mulheres (52.09%)
- Diferença pequena, sugerindo que gênero não é um fator determinante.

3.4.3 Educação vs. Diabetes

Pessoas com menor escolaridade têm mais diabetes:

- **Sem estudo:** 0.13% dos diabéticos vs. 0.08% dos não diabéticos.
- **Ensino primário:** 3.35% dos diabéticos vs. 1.31% dos não diabéticos.
- **Ensino superior:** 29.42% dos diabéticos vs. 44.19% dos não diabéticos.

Maior nível de escolaridade pode estar associado a um melhor acesso à informação sobre saúde e a hábitos mais saudáveis. Além disso, a escolaridade está relacionada a uma renda maior, o que proporciona um melhor acesso aos serviços de saúde.

3.4.4 Renda vs. Diabetes

Quanto menor a renda, maior a incidência de diabetes:

- **6.74% dos diabéticos ganham menos de \$10.000**, contra 3.47% dos não diabéticos.
- **Apenas 20.36% dos diabéticos têm renda superior a \$75.000, enquanto entre os não diabéticos esse valor é 38.06%.**

Pessoas de baixa renda podem ter menos acesso a alimentação saudável e cuidados preventivos de saúde.

4 Correlações e seus Significados

4.1 Saúde Geral vs. Diabetes (0.414), Pressão Alta (0.324), Problema Cardíaco (0.268), Dificuldade de Andar (0.466)

- Pessoas com pior saúde geral tendem a ter essas condições crônicas, indicando que esses fatores estão interligados.

4.2 Dificuldade de Andar vs. Saúde Geral (0.466), Diabetes (0.272), Pressão Alta (0.234), Problema Cardíaco (0.232)

- Confirma que doenças crônicas impactam a mobilidade.

4.3 Ensino vs. Faixa Renda (0.463)

- Quanto maior o nível de ensino, maior a renda, reforçando o impacto da educação na condição financeira.

4.4 Faixa Renda vs. Saúde Geral (-0.382), Dificuldade de Andar (-0.336), Diabetes (-0.232)

- Pessoas com menor renda tendem a ter pior saúde, provavelmente por terem acesso limitado a serviços de saúde, alimentação de qualidade, locais para praticarem atividades físicas.

4.5 Hábitos e Doenças

4.5.1 Atividade Física vs. Saúde Geral (-0.270), Dificuldade de Andar (-0.276)

- Pessoas mais ativas tendem a relatar melhor saúde e menos dificuldades de locomoção.

4.5.2 Atividade Física vs. Diabetes (-0.158), Pressão Alta (-0.136)

- Atividade física parece reduzir a probabilidade dessas condições, o que está de acordo com o esperado.

4.5.3 Fumar vs. Problema Cardíaco (0.124), Diabetes (0.085), Pressão Alta (0.087)

- Reforça que o tabagismo pode estar associado a problemas cardiovasculares e metabólicos.

4.6 Acesso à Saúde e Condições Financeiras

4.6.1 Sem Dinheiro para Consultas vs. Plano de Saúde (-0.221), Faixa de Renda (-0.195), Saúde Geral (0.166)

- Pessoas sem plano de saúde e com menor renda relatam pior estado de saúde, sugerindo barreiras no acesso à saúde.