



SPRINT 03

DATA SCIENCE & STATISTICAL COMPUTING

DANIEL DE SOUZA CARVALHO

LepIn

(Laparoscopy Inclusive)

JULIO CESAR ZAMPIERI RM98772

KAYKY OLIVEIRA SHUNCK RM99756

LEONARDO SHUNCK RAINHA RM99902

LUCAS CARLOS BANDEIRA TEIXEIRA RM98640

JOÃO GABRIEL DIAS DE MELLO DO NASCIMENTO RM99092

Resumo

Este projeto analisa um conjunto de dados sobre casos de diabetes, explorando a relação entre diversas variáveis de saúde, como pressão alta, colesterol, IMC (Índice de Massa Corporal), tabagismo e atividade física. O objetivo é identificar padrões e correlações entre essas variáveis e a presença de diabetes, além de realizar testes de hipótese para verificar a significância dessas associações.

Tratamentos Realizados na Base de Dados

1. Leitura e Limpeza dos Dados:

- 1.1 Carregamento do arquivo CSV contendo as variáveis relevantes para o estudo.
- 1.2 Verificação e tratamento de valores ausentes.

2. Visualizações gráficas:

- **Foram criados quatro gráficos principais para explorar relações importantes no conjunto de dados:**

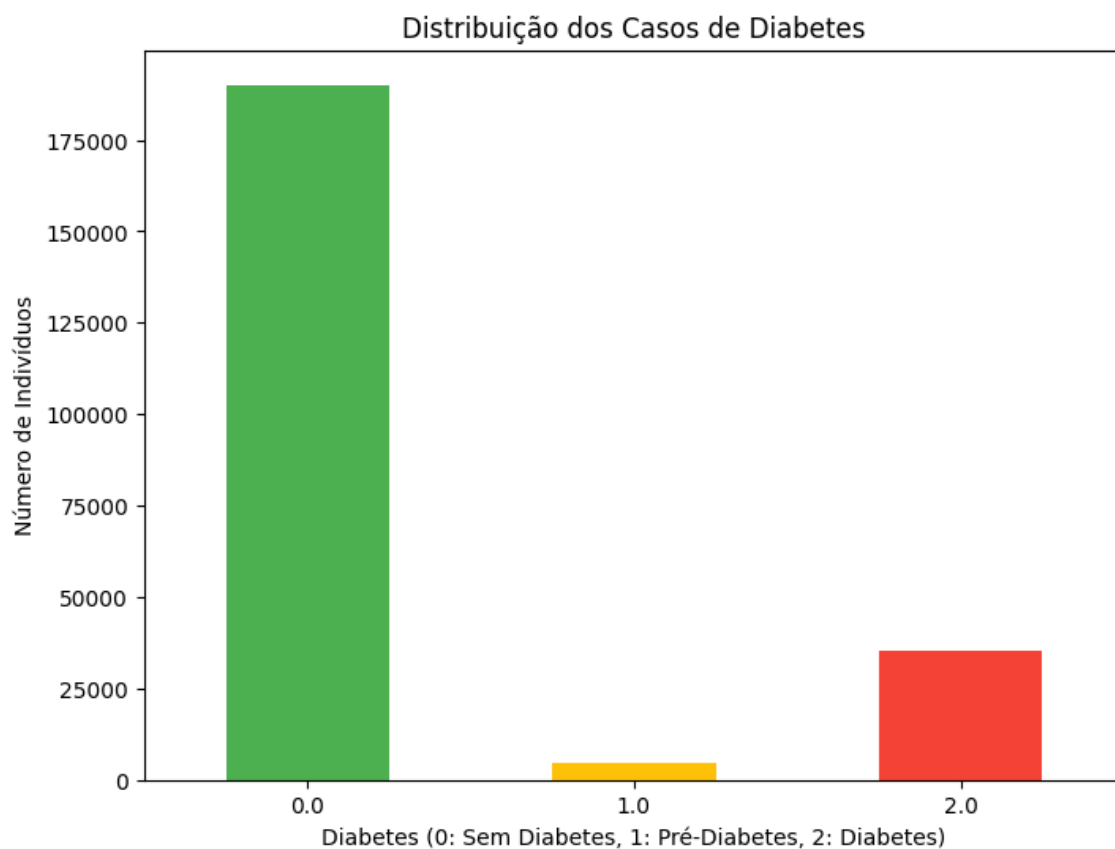
- 2.1 Distribuição dos casos de diabetes.
- 2.2 Comparação entre fumantes e casos de diabetes.
- 2.3 Relação entre atividade física e diabetes.
- 2.4 Relação entre pressão alta e diabetes.
- 2.5 Comparação entre homens e mulheres em relação aos casos de diabetes.

Estatísticas Descritivas Gerais

- **Diabetes_012:** (0 = sem diabetes, 1 = pré-diabetes, 2 = diabetes)
- **HighBP:** (0 = sem pressão alta, 1 = pressão alta)
- **HighChol:** (0 = sem colesterol alto, 1 = colesterol alto)
- **BMI:** Índice de Massa Corporal (contínuo)
 - **Média:** ~27,5
 - **Intervalo de valores entre pessoas com peso normal até obesidade**
- **Smoker:** (0 = não, 1 = sim)
- **PhysActivity:** Atividade física nos últimos 30 dias (0 = não, 1 = sim)
 - **0 dias:** Indica que a pessoa não teve nenhum problema de saúde física no último mês.
 - **30 dias:** Significa que a pessoa teve problemas físicos durante todos os dias do último mês.
- **GenHlth:** Saúde geral autoavaliada numa escala de 1 a 5 (1 = excelente, 5 = ruim)
- **Teste de Hipótese:** Para o teste de hipótese, adotaremos (H_0) para **hipótese nula** e (H_1) para **hipótese alternativa**.

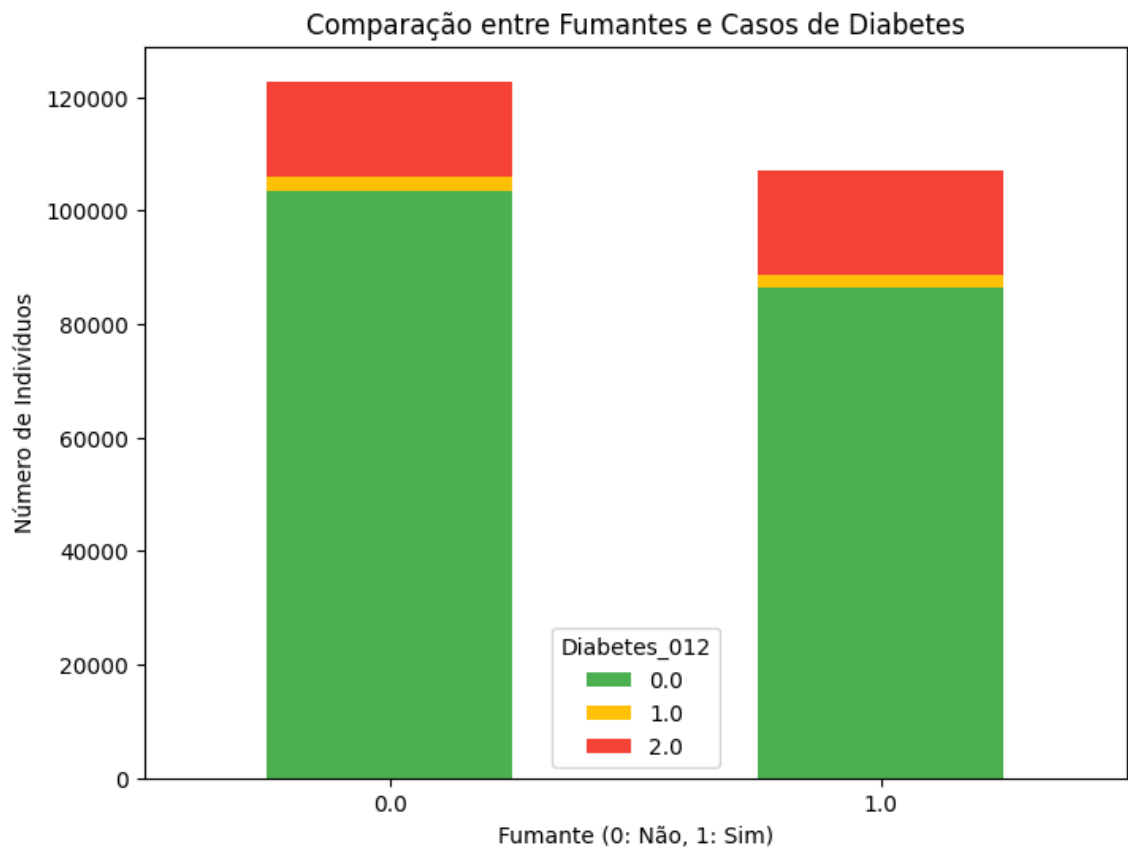
Leitura dos Gráficos

1. Gráfico de Barra – Número de casos (Apresentação)



1.1 – O gráfico apresenta o número de casos de diabetes.

2. Gráfico de Barra – Comparação entre fumantes e não fumantes.



2.1 - Este gráfico mostra a distribuição dos casos de diabetes (sem diabetes, pré-diabetes e diabetes) entre fumantes e não fumantes. Cada barra é dividida para representar a prevalência de diabetes em ambos os grupos.

2.2 - Observa-se que a proporção de indivíduos com diabetes é **maior entre fumantes**, enquanto a maioria dos não fumantes se enquadra na categoria de **sem diabetes**.

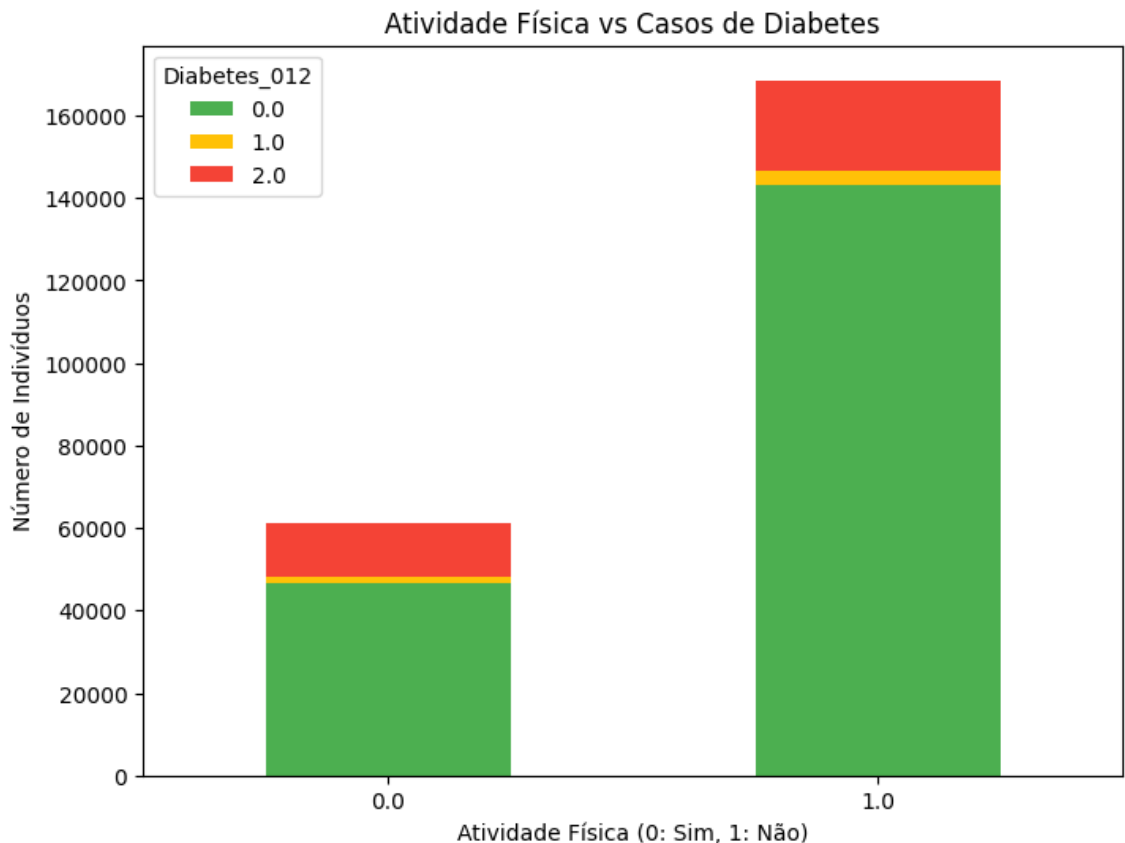
2.3 - Este padrão sugere uma **associação entre fumar e ter diabetes**, com os fumantes apresentando uma maior probabilidade de desenvolver a doença.

2.4 – Teste de hipótese:

2.4.1 - **P-valor:** 5.16×10^{-110}

2.4.2 - Com base nos resultados, podemos concluir que há uma **associação significativa entre fumar e ter diabetes**. Ou seja, a prevalência de diabetes é significativamente maior entre fumantes do que entre não fumantes.

3. Gráfico de Barra – Atividades físicas vs casos de diabetes.



3.1 - Este gráfico mostra a distribuição dos casos de diabetes (sem diabetes, pré-diabetes e diabetes) entre quem pratica atividades físicas e sedentários. Cada barra é dividida para representar a prevalência de diabetes em ambos os grupos.

3.2 - Observa-se que a proporção de indivíduos com diabetes é **maior entre sedentários**, enquanto a maioria dos desportistas se enquadra na categoria de **sem diabetes**.

3.3 - Este padrão sugere uma **associação entre praticar ou não atividades físicas**, com os sedentários apresentando uma maior probabilidade de desenvolver a doença.

3.4 – Teste de hipótese:

3.4.1 – P-valor: 5.16×10^{-123}

3.4.2 - Com base nos resultados, podemos concluir que há uma **associação significativa entre ser ativo fisicamente e ter diabetes**. Ou seja, a prevalência de diabetes é significativamente maior entre quem não pratica atividades.

