

Data Mining e Graph Mining

Inteligência Artificial

Professor André Luiz Esperidião

LABORATÓRIO

Nesta apresentação, exploraremos a aplicação da metodologia CRISP-DM para identificar e mitigar a perda de clientes em uma empresa de telecomunicações. Nosso objetivo é transformar dados brutos em insights açãoáveis.



Nossa abordagem: CRISP-DM em 6 Fases

METODOLOGIA CRISP-DM

Seguiremos a estrutura do CRISP-DM, analisando cada fase sob três perspectivas cruciais: o que o negócio almeja, o papel do cientista de dados e o resultado concreto esperado.

1

Business Understanding

Definição clara dos objetivos de negócio.

2

Data Understanding

Exploração e compreensão dos dados disponíveis.

3

Data Preparation

Limpeza e transformação dos dados para modelagem.

4

Modeling

Construção e seleção de modelos preditivos.

5

Evaluation

Avaliação do desempenho dos modelos.

6

Deployment

Implementação da solução e acompanhamento.

CRISP-DM

FASE 02

Data Understanding: Desvendando Nossos Dados

A fase de compreensão dos dados nos permite uma visão 360° do cliente, integrando perfil, contrato, uso do serviço, experiência e comportamento financeiro.



Histórico de Uso

Serviços utilizados e padrões de consumo.



Reclamações e Atendimento

Registros de interações com o suporte técnico.



Perfil Demográfico

Gênero, idade (idoso), parceiro, dependentes.



Dados de Faturamento

Valores mensais e totais, métodos de pagamento.

CRISP-DM

Estrutura do Dataset: Visão Geral

Nosso dataset possui **6000 linhas e 27 colunas**, oferecendo uma riqueza de informações para análise.

Identificação	customer_id	Identificador único do cliente
Demográfico	gender, senior_citizen, partner, dependents	Informações de perfil do cliente
Relacionamento	tenure_months	Tempo de relacionamento em meses
Serviços	phone_service, internet_service, etc.	Detalhes dos serviços contratados
Contrato/Cobrança	contract, monthly_charges, payment_method	Condições de contrato e faturamento
Qualidade	complaints_last_3m, outages_last_3m	Registros de problemas recentes
Financeiro	late_payments_last_6m	Histórico de pagamentos em atraso
Variável Alvo	churn	Indica se o cliente cancelou (1) ou não (0)

CRISP-DM

Análise Descritiva dos Dados

- **Count:** Quantidade de valores não nulos.
- **Mean:** Média aritmética, sensível a outliers.
- **Std:** Desvio padrão, mede a dispersão dos dados.
- **Min/Max:** Valores mínimo e máximo, identificando limites.
- **Quartis (25%, 50%, 75%):** Indicadores de distribuição central.

CRISP-DM

Análise da Taxa de Churn (Base Rate)

A proporção de clientes que cancelaram o serviço é um indicador crítico da saúde do negócio.

Clientes Ativos

85.95%

Dos 6000 clientes, **5157** permanecem utilizando os serviços.

Clientes em Churn

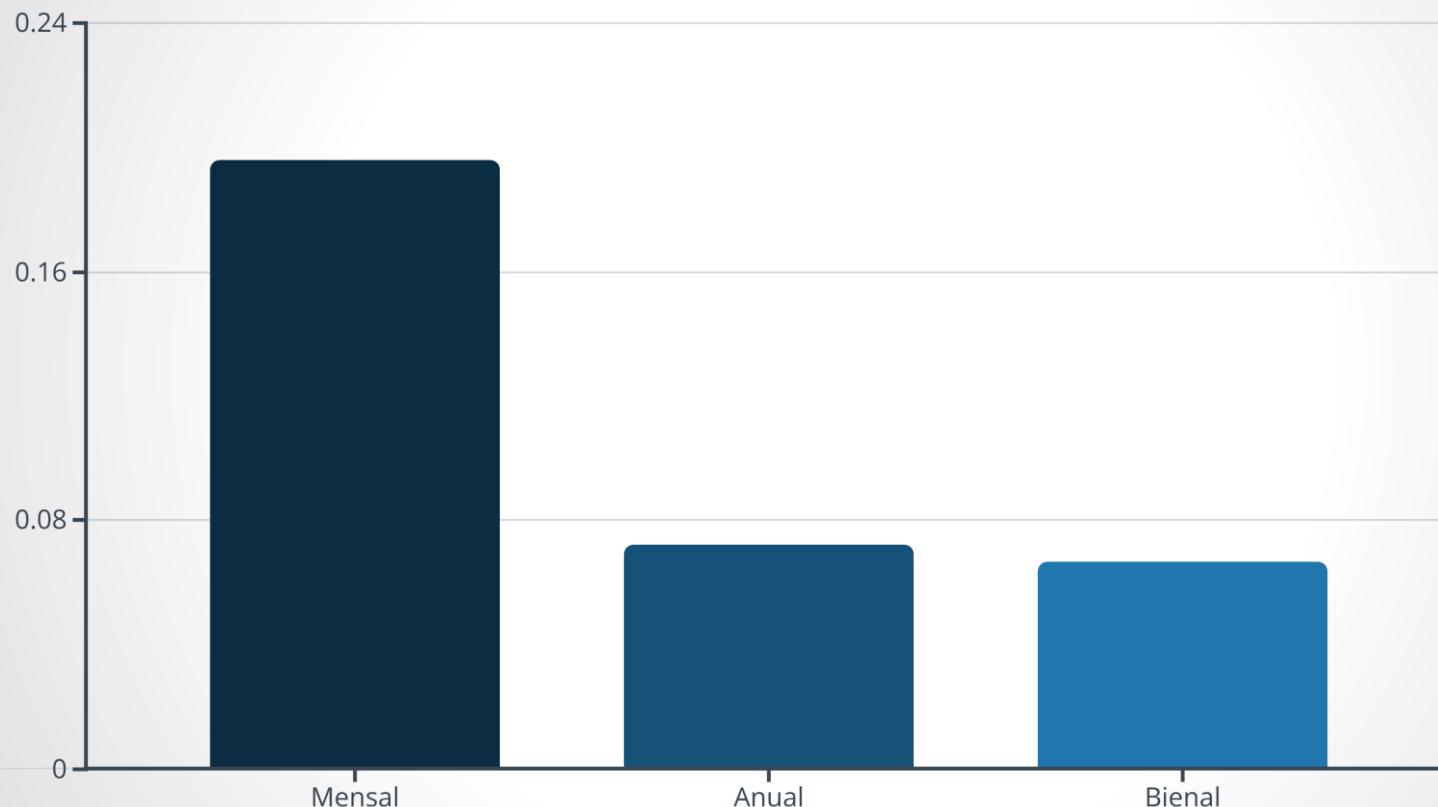
14.05%

Uma parcela de **843** clientes optou por cancelar seus serviços, representando nossa variável alvo.

Observação didática: Este é um cenário real de desbalanceamento de classes, comum em problemas de churn, e que exige atenção especial na modelagem.

CRISP-DM

Churn por Tipo de Contrato: Um Fator Chave



O contrato **mensal** (Month-to-month) apresenta a maior taxa de churn, quase **três vezes maior** que os contratos anuais e bienais.

Insight de Negócio: Foco nos Contratos Mensais

Os dados revelam uma oportunidade clara para a empresa reduzir o churn através de estratégias de retenção.

Alto Risco no Mensal

Contratos mensais são o principal fator de risco de churn, exigindo atenção prioritária.

Migração Estratégica

Migrar clientes de contratos mensais para anuais pode reduzir o churn em mais de 60%.

Incentivos Direcionados

Ofertas personalizadas para clientes mensais com sinais de risco podem impulsionar a conversão para contratos de maior prazo.

Ganho Marginal do Bienal

O contrato bienal oferece uma redução marginal em relação ao anual, questionando a necessidade de incentivos agressivos.

Conexão com Data Mining: A Força da Feature 'Contract'

A variável 'contract' é uma "feature" (característica) **fortemente preditiva** e uma excelente candidata para diversas aplicações no modelo.



Tratá-la como uma **feature ordinal** (Month-to-month < One year < Two year) permitirá que o modelo capture a hierarquia e o risco associado a cada tipo de contrato, otimizando a precisão das previsões de churn.

CRISP-DM

Próximo passo natural é cruzar churn com:

- Tempo de contrato (Tenure)
- Número de reclamações
- Pagamento e Inadimplência



CRISP-DM

Churn vs Tempo de Contrato (Tenure):

- Medidas descritivas de não-churn e churn por tempo de contrato

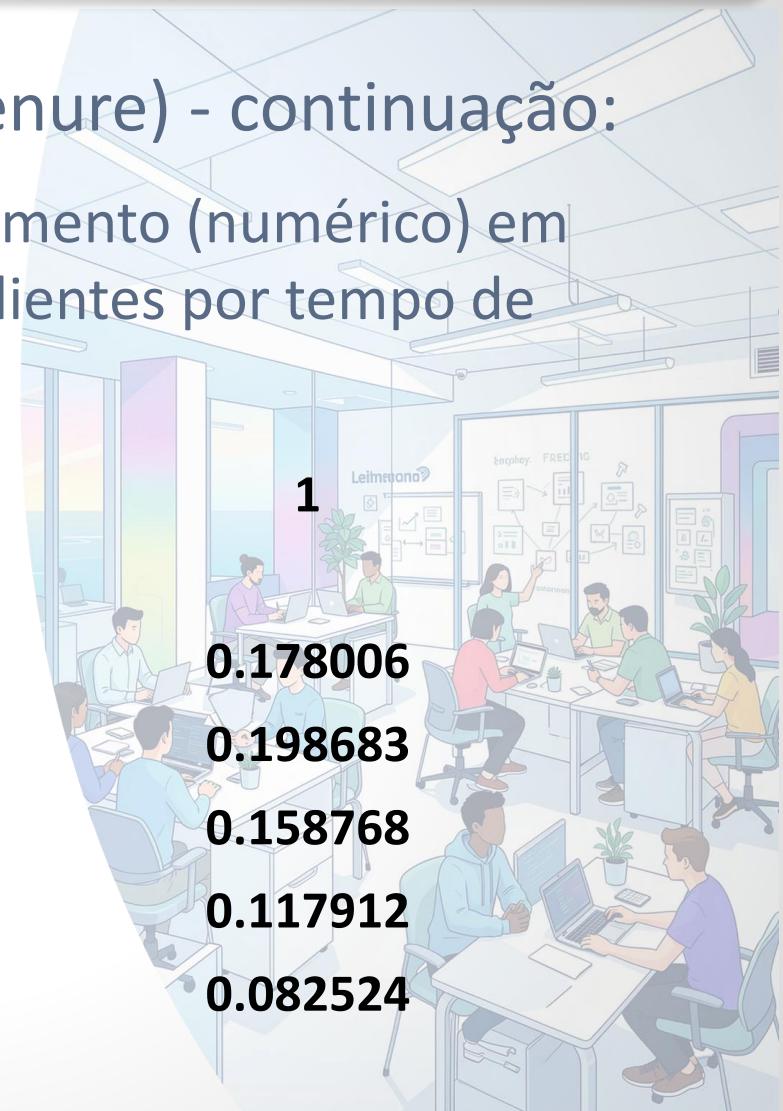
churn	count	mean	std	min	25%	50%	75%	max
0	5157	29.832	22.464	0.0	8.0	27.0	49.0	72.0
1	843	21.668	19.938	0.0	6.0	13.0	35.0	72.0

CRISP-DM

Churn vs Tempo de Contrato (Tenure) - continuação:

- Convertendo o tempo de relacionamento (numérico) em faixas categóricas para “agrupar” clientes por tempo de contrato.

Churn	0	1
Tenure group		
0-6	0.821994	0.178006
6-12	0.801317	0.198683
12-24	0.841232	0.158768
24-48	0.882088	0.117912
48+	0.917476	0.082524



Churn vs Tempo de Contrato (Tenure) - continuação:

📌 Insight:

- Churn concentrado nos primeiros meses
- Onboarding ruim = churn alto (telecom sofre disso há décadas e finge surpresa)



CRISP-DM

Churn vs Número de Reclamações:

- Média de não-churn e churn com reclamações nos últimos 3 meses

churn	mean
0	0.373667
1	0.623962

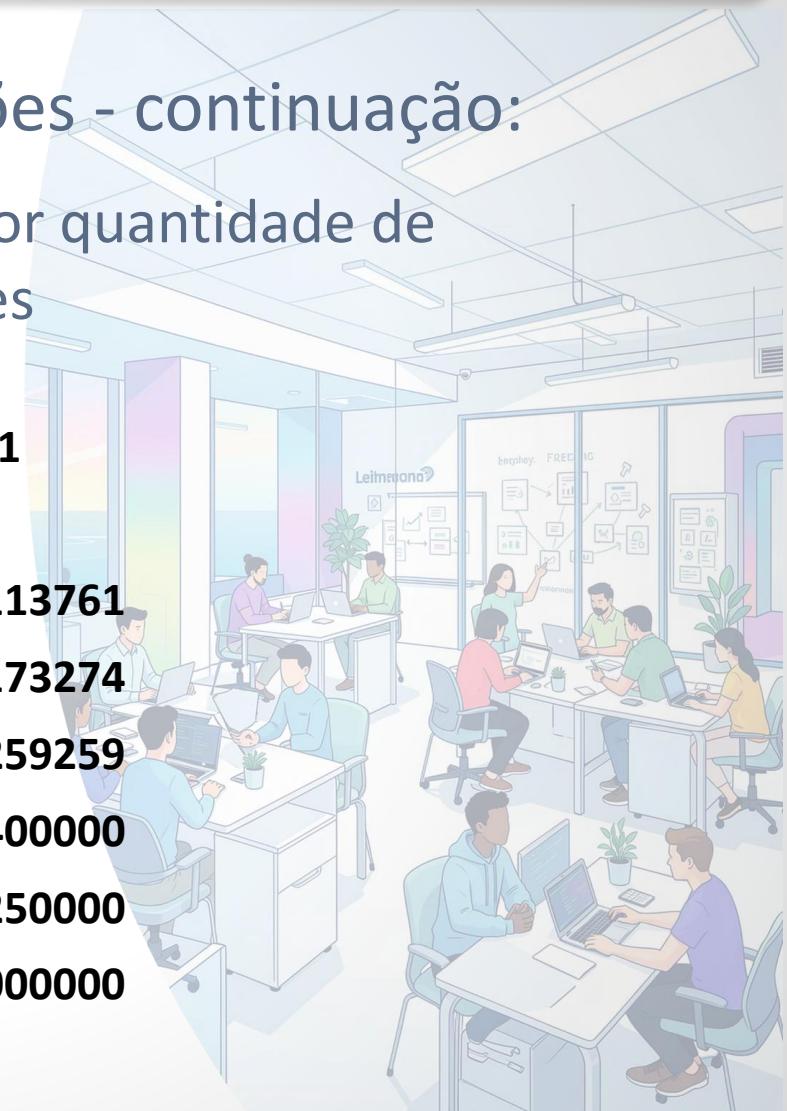


CRISP-DM

Churn vs Número de Reclamações - continuação:

- Proporção de não-churn e churn por quantidade de reclamações nos últimos três meses

Churn	0	1
complaints_last_3m		
0	0.886239	0.113761
1	0.826726	0.173274
2	0.740741	0.259259
3	0.600000	0.400000
4	0.750000	0.250000
5	0.000000	1.000000



Churn vs Número de Reclamações - continuação:

- 📌 Insight – Lei universal do churn:
 - Cliente que reclama muito → vai embora



CRISP-DM

Pagamento e Inadimplência:

- Média de não-churn e churn cliente com pagamento em atraso nos últimos 6 meses

churn	mean
0	0.375024
1	0.478055



CRISP-DM

Como interpretar isso corretamente?

- ◆ Clientes que não churnaram (churn = 0)

- Média $\approx 0,38$ atrasos

- ◆ Clientes que churnaram (churn = 1)

- Média $\approx 0,48$ atrasos

👉 Em média, quem churnou atrasou mais pagamentos

⚠️ Mas atenção: a diferença é pequena

- Diferença absoluta: $0.478 - 0.375 \approx 0.103$

📌 Isso significa:

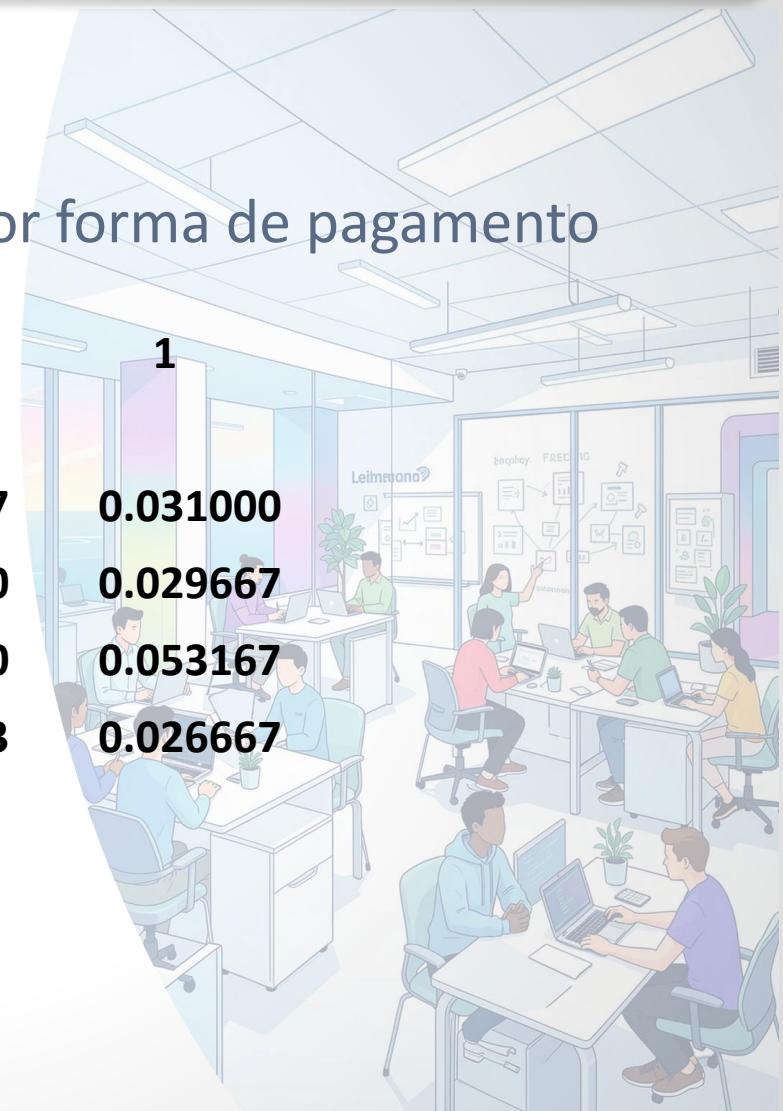
- Existe relação, mas não é forte isoladamente
 - Esse NÃO é um “killer feature” sozinho



Churn vs Forma de Pagamento

- Proporção de não-churn e churn por forma de pagamento

Churn	0	1
payment_method		
Bank transfer (automatic)	0.223167	0.031000
Credit card (automatic)	0.203500	0.029667
Electronic check	0.276500	0.053167
Mailed check	0.156333	0.026667



Churn vs Forma de Pagamento

- Média de churn por forma de pagamento

payment_method	mean
Bank transfer (automatic)	0.121967
Credit card (automatic)	0.127234
Electronic check	0.161274
Mailed check	0.145719



Churn vs Forma de Pagamento:

📌 Insights

- Electronic check = maior churn
- Atraso é proxy de insatisfação + risco financeiro



Conclusão da Fase Data Understanding

- ✓ Já sabemos que churn está fortemente ligado a:
 - Tipo de contrato
 - Tempo de contrato (tenure)
 - Reclamações
 - Atrasos
 - Tipo de pagamento



Data Mining e Graph Mining

Inteligência Artificial

Professor André Luiz Esperidião