

Sumário

1	Introdução	2
2	Ferramentas Utilizadas	2
3	Análises Univariadas	3
3.1	Distribuição do Churn	3
3.2	Distribuição de Tempo de Contrato (Tenure)	4
3.3	Distribuição de Cobrança Total (TotalCharges)	5
3.4	Distribuição de Cobrança Mensal (MonthlyCharges)	6
4	Grupo 1 - Churn vs (Gender SeniorCitizen Partner Dependents)	7
4.1	Churn vs Gênero	7
4.2	Churn vs Idosos (SeniorCitizen)	8
4.3	Churn vs Parceiros (Partner)	8
4.4	Churn vs Dependentes	9
5	Grupo 2 - Churn vs (PhoneService MultipleLines InternetService StreamingTV StreamingMovies)	10
5.1	Churn vs Serviço de telefonia (PhoneService)	10
5.2	Churn vs Múltiplas Linhas (MultipleLines)	11
5.3	Churn vs Serviços de internet (InternetService)	11
5.4	Churn vs Streaming de TV	12
5.5	Churn vs Streaming de Filmes	13
6	Grupo 3 - Churn vs (OnlineSecurity OnlineBackup DeviceProtection TechSupport)	14
6.1	Churn vs Serviço de Segurança Online (OnlineSecurity)	14
6.2	Churn vs Serviço de Backup (OnlineBackup)	15
6.3	Churn vs Serviço de Proteção para Dispositivos (DeviceProtection)	16
6.4	Churn vs Suporte Técnico (TechSupport)	17
7	Grupo 4 - Churn vs (Contract PaperlessBilling PaymentMethod)	18
7.1	Churn vs Método de Pagamento	18
7.2	Churn por Tipo de Contrato	19
8	Perfis de Risco	19

Análise Exploratória de Dados (EDA) - Telecom Churn

Reyso Teixeira

1 Introdução

O problema de *Customer Churn*, ou evasão de clientes, é um desafio relevante no setor de telecomunicações, caracterizado pela alta competitividade e pela facilidade com que os clientes podem migrar entre provedores. O conjunto de dados **Telco Customer Churn** disponibilizado no Kaggle apresenta informações de aproximadamente 7.043 clientes de uma empresa fictícia de telecomunicações que oferece serviços de telefonia fixa, Internet e pacotes adicionais, como streaming e suporte técnico.

O objetivo principal é **prever quais clientes estão propensos a cancelar o serviço (churn)**, com base em variáveis demográficas, financeiras e relacionadas aos serviços contratados. Esse tipo de análise é essencial, pois o custo para reter um cliente existente costuma ser significativamente menor do que o custo para adquirir novos clientes.

Entre as variáveis analisadas destacam-se:

- **Dados demográficos:** gênero, idade (*SeniorCitizen*), status de parceiro(a) e dependentes.
- **Serviços contratados:** tipos de Internet (DSL, fibra óptica), serviços de telefonia, streaming, segurança online, backup, proteção de dispositivos e suporte técnico.
- **Características financeiras:** cobranças mensais (*MonthlyCharges*), cobranças acumuladas (*TotalCharges*), tipos de contrato e métodos de pagamento.

A análise do churn permite entender os fatores de maior impacto na decisão de cancelamento e, conseqüentemente, direcionar ações de retenção mais eficientes. Identificar perfis de clientes com maior propensão ao churn possibilita à empresa desenvolver estratégias proativas para minimizar perdas e maximizar a fidelização.

2 Ferramentas Utilizadas

Para a realização desta análise, foram utilizadas as seguintes ferramentas e tecnologias:

- **Databricks:** Plataforma de análise de dados utilizada para a execução de notebooks, organização dos dados e processamento das tabelas. A integração com o ambiente distribuído permitiu a ingestão e análise dos dados de forma rápida e escalável.

- **Python:** Linguagem de programação utilizada para análise exploratória dos dados (EDA), criação de gráficos, manipulação de tabelas e construção de modelos. Bibliotecas como *pandas*, *matplotlib* e *seaborn* foram amplamente empregadas.
 - **SQL:** Linguagem utilizada no ambiente Databricks para consultas diretas às tabelas, filtragem e agregação de informações essenciais para a análise.
 - **LaTeX:** Ferramenta empregada para a elaboração deste documento técnico, visando garantir uma apresentação formal e bem estruturada dos resultados.
-

3 Análises Univariadas

3.1 Distribuição do Churn

Percentual e Quantidade de Churn na Base de Dados

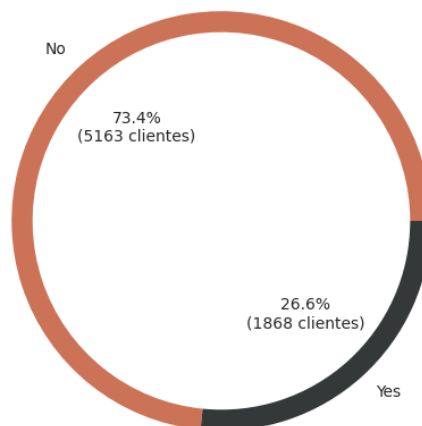


Figura 1: Distribuição do Churn

Aproximadamente 26.6% dos clientes cancelaram o serviço. Aqui apresenta-se um desbalanceamento na quantidade da amostra de usuários. A quantidade de cancelamentos é bem menor que a quantidade de não cancelamentos.

3.2 Distribuição de Tempo de Contrato (Tenure)

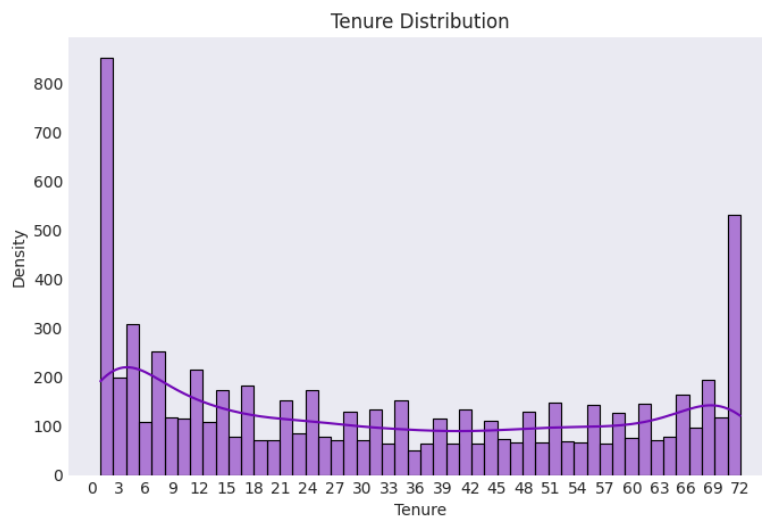


Figura 2: Distribuição do tempo de contrato

Analisando a frequência dos valores de tempo de contrato, a figura sugere uma distribuição bimodal. Os maiores números de tempo de contrato concentram-se nos primeiros e nos últimos 6 meses deste recorte de tempo da base de dados.

No entanto, observa-se que os clientes com menos tempo de contrato são a maioria. Precisamos investigar se esses clientes também representam um maior número de cancelamentos (Churn).

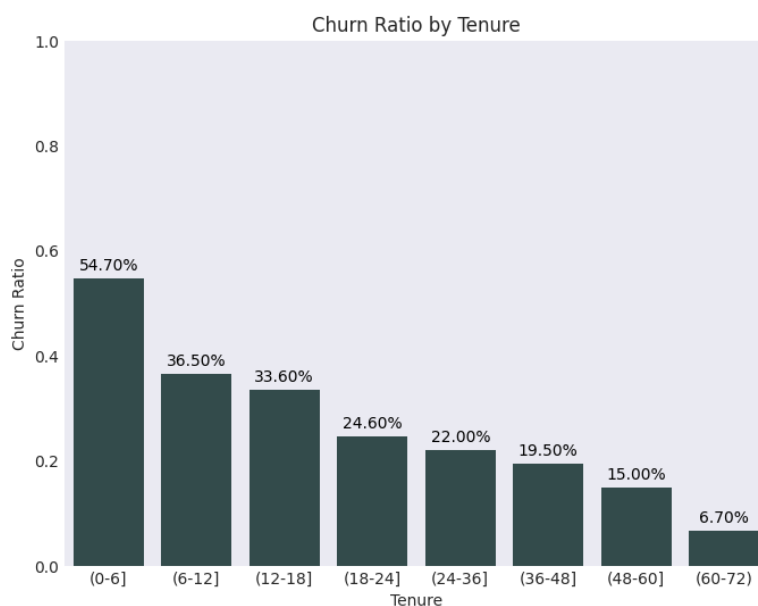


Figura 3: Proporção de churn por tempo de contrato

De fato, mais de 50% dos clientes que possuem tempo de contrato de 0 a 6 meses

cancelaram o serviço. Isso destaca que um menor tempo de contrato é um elemento importantíssimo para construir um perfil de risco.

3.3 Distribuição de Cobrança Total (TotalCharges)

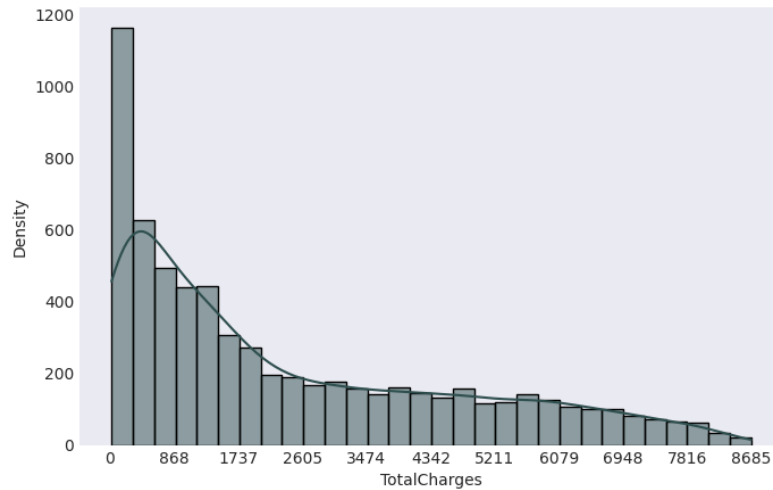


Figura 4: Distribuição de TotalCharges

A distribuição é assimétrica, concentrada em clientes com baixas cobranças acumuladas.

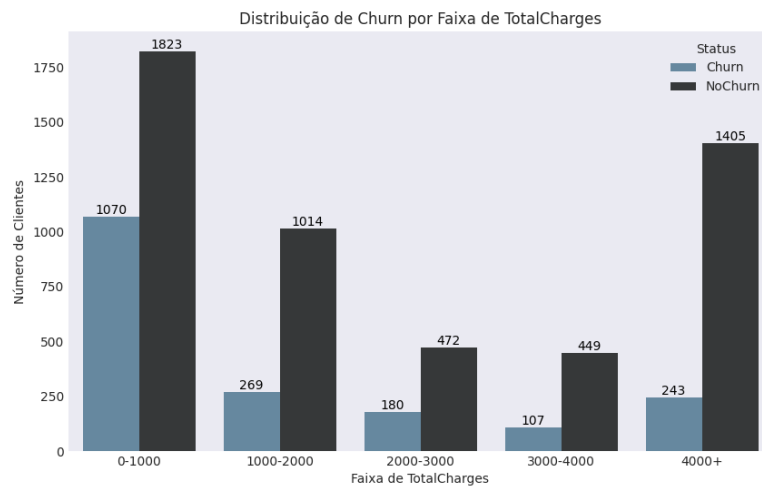


Figura 5: Proporção de churn por TotalCharges segmentado

Segmentando o total gasto pelo cliente (TotalCharges) em 4 intervalos, observamos que o maior número de cancelamento de serviço ocorre na faixa de menor valor de até \$1000. O que é perfeitamente compreensível, uma vez que clientes com menos tempo de contrato tendem a cancelar o serviço e o montante arrecadado pela empresa tende a ser menor.

3.4 Distribuição de Cobrança Mensal (MonthlyCharges)

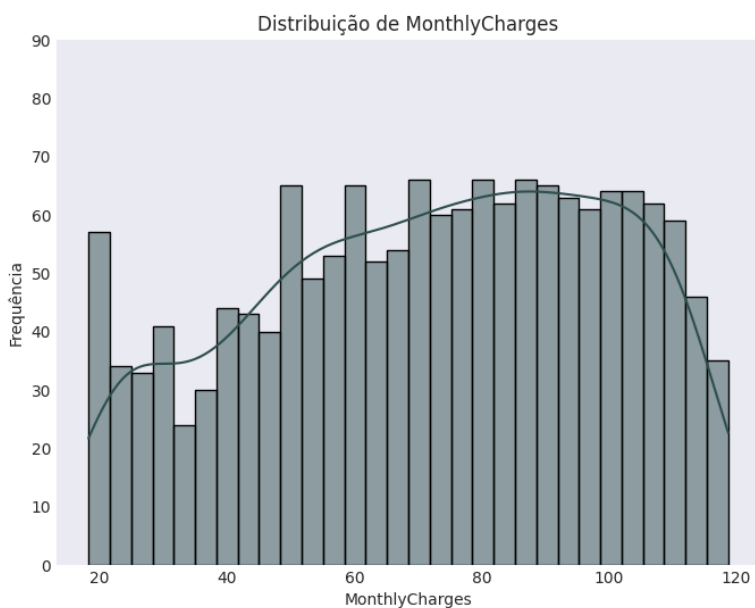


Figura 6: Distribuição de MonthlyCharges

Muitos clientes pagam tarifas mensais moderadas, mas existem planos com tarifas mais elevadas.

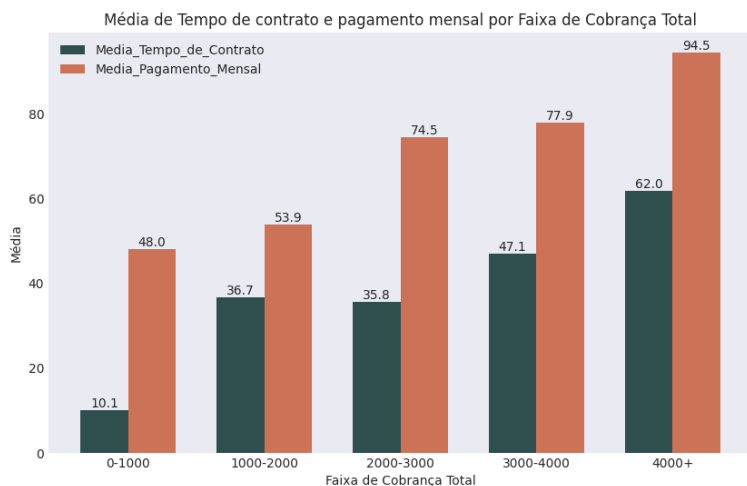


Figura 7: Média de tempo de contrato e cobrança mensal por faixa de gasto total

Apesar dos clientes da faixa de gastos totais até \$1000 serem maioria na base de dados analisada, eles possuem as menores médias de tempo de contrato e de pagamento mensal, cerca de 10.1 meses e \$48/mês, respectivamente.

Ao contrário dos clientes da faixa \$4000+, que possuem as maiores médias de tempo de contrato e de pagamento mensal. Além disso, aproximadamente 85% desses clientes não cancelaram o serviço.

4 Grupo 1 - Churn vs (Gender | SeniorCitizen | Partner | Dependents)

4.1 Churn vs Gênero

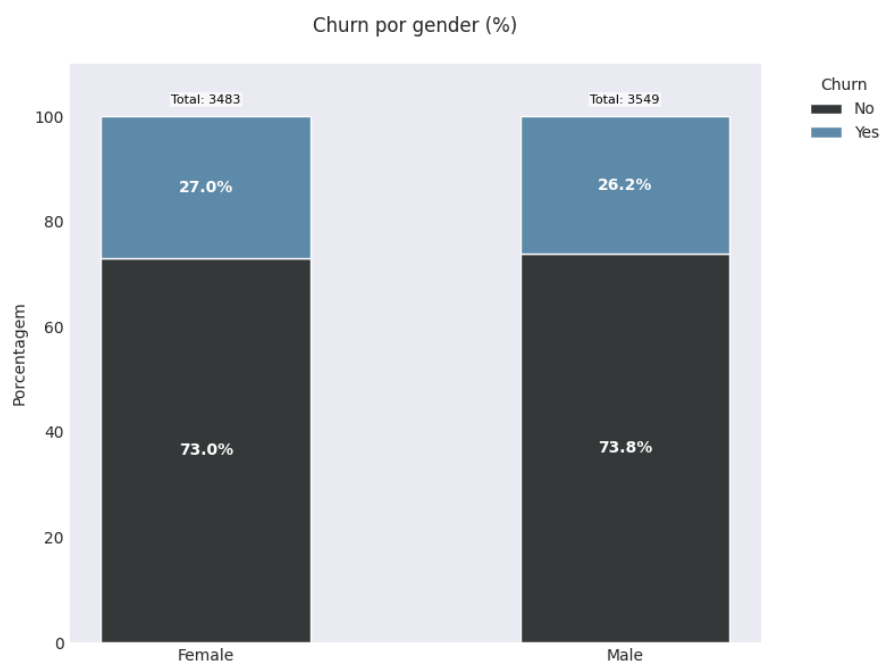


Figura 8: Churn por gênero

Há um balanceamento na amostra de clientes do sexo feminino e masculino.

4.2 Churn vs Idosos (SeniorCitizen)

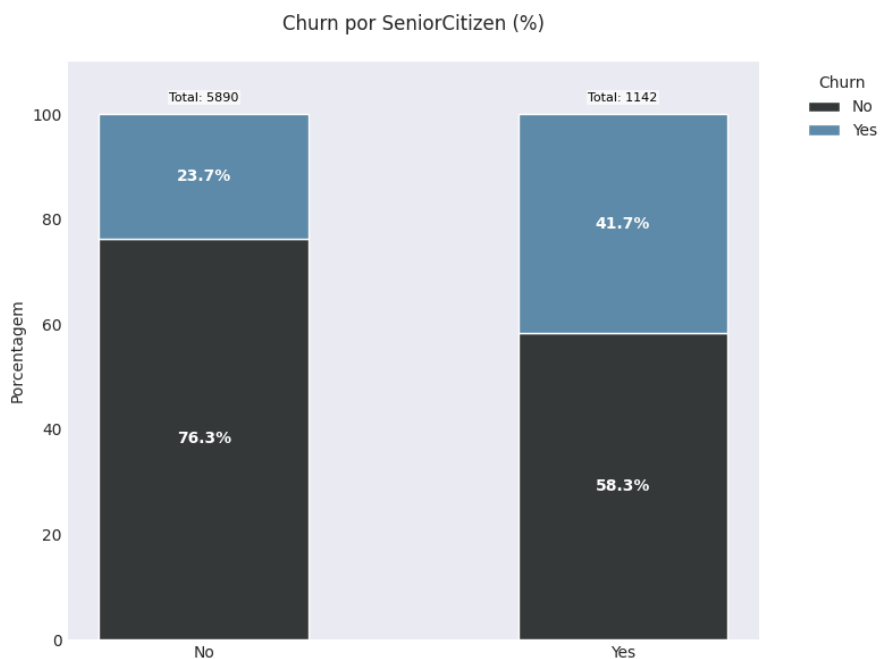


Figura 9: Churn por SeniorCitizen

Temos um número muito pequeno de clientes idosos, cerca de 1142. No entanto, cerca de 41% deles cancelaram o serviço com a empresa, o que é uma porcentagem significativa diante do número total desse público.

4.3 Churn vs Parceiros (Partner)

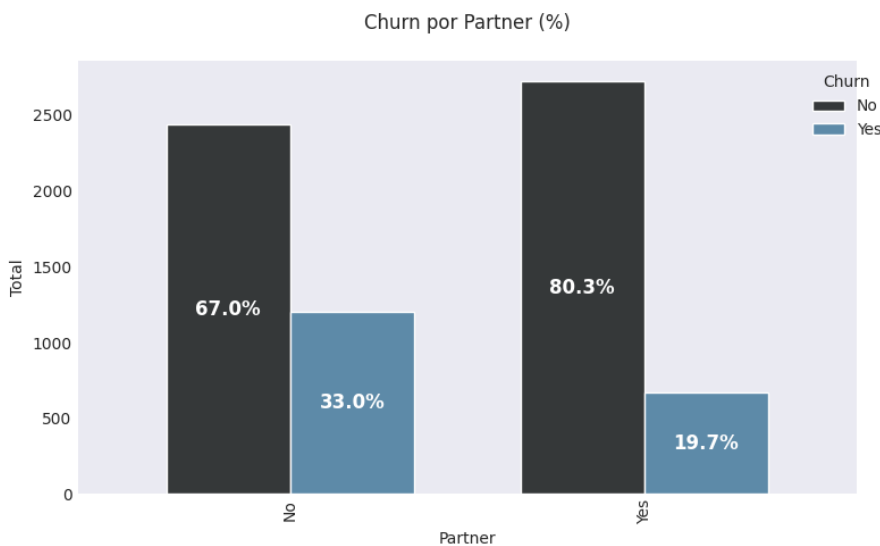


Figura 10: Churn por Partner

Clientes que possuem parceiros(as) têm menos cancelamentos em comparação com aqueles que vivem sozinhos.

4.4 Churn vs Dependentes

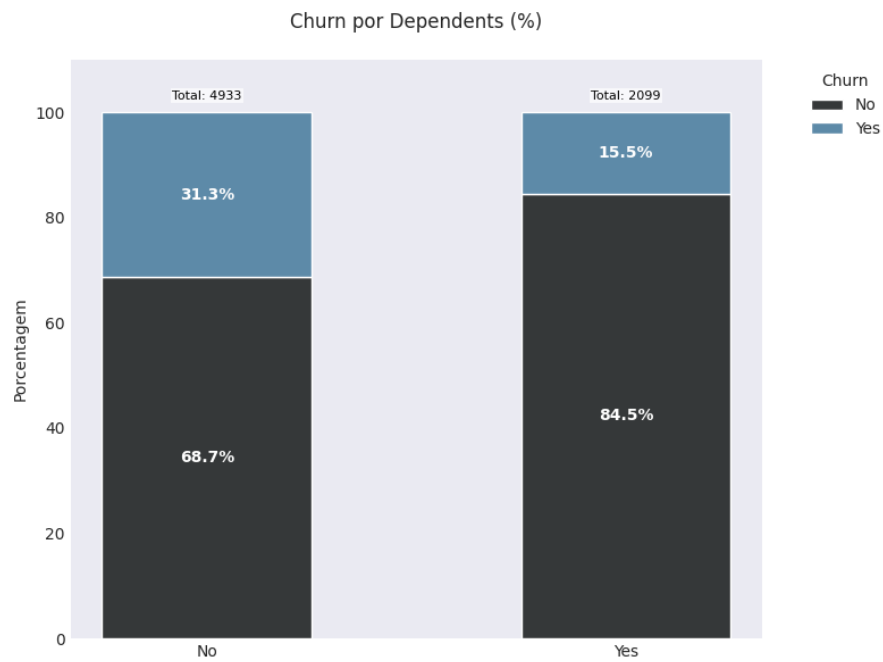


Figura 11: Churn por Dependents

Semelhante ao caso dos parceiros, os casos de cancelamento são maiores para aqueles clientes que não têm dependentes.

5 Grupo 2 - Churn vs (PhoneService | MultipleLines | InternetService | StreamingTV | StreamingMovies)

5.1 Churn vs Serviço de telefonia (PhoneService)

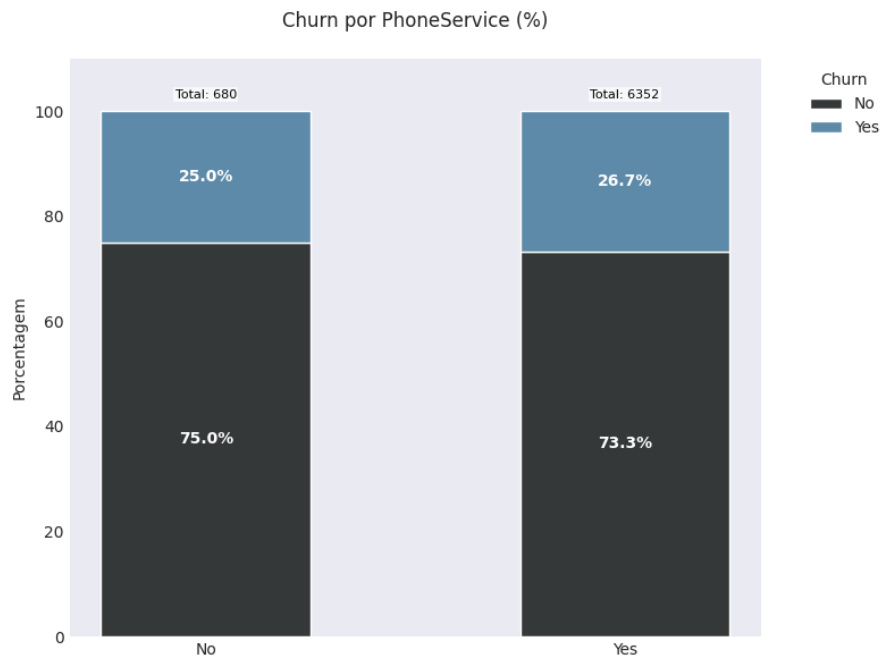


Figura 12: Churn por PhoneService

Um número expressivo de clientes, 90% (6352 clientes), possui serviço de telefonia. Destes clientes, 26.7% cancelaram o serviço por algum motivo.

5.2 Churn vs Múltiplas Linhas (MultipleLines)

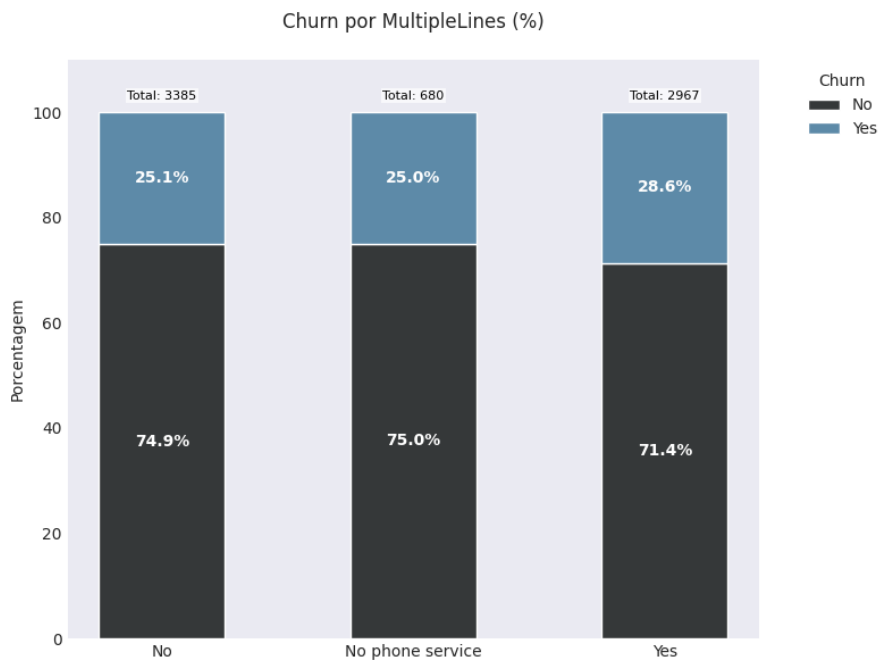


Figura 13: Churn por MultipleLines

Cerca de 2667 dos clientes possuem múltiplas linhas disponíveis em suas residências. É interessante notar que ter ou não múltiplas linhas apresenta um número muito semelhante de cancelamento, o que sugere que isto pode não ser um fator determinante para o rompimento de contrato.

5.3 Churn vs Serviços de internet (InternetService)

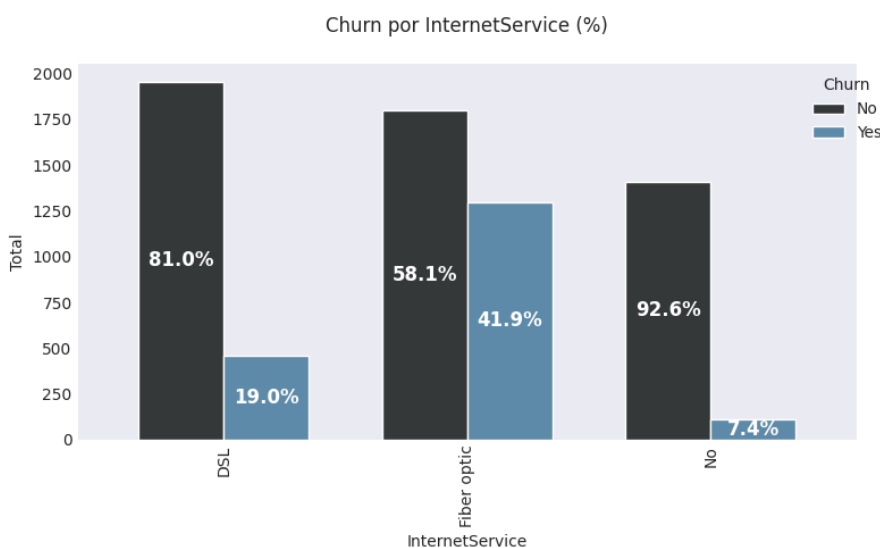


Figura 14: Churn por InternetService

Entre os 3096 clientes que assinam o serviço de internet por fibra óptica, observamos um elevado número de cancelamentos, cerca de 42% (1297 clientes).

5.4 Churn vs Streaming de TV

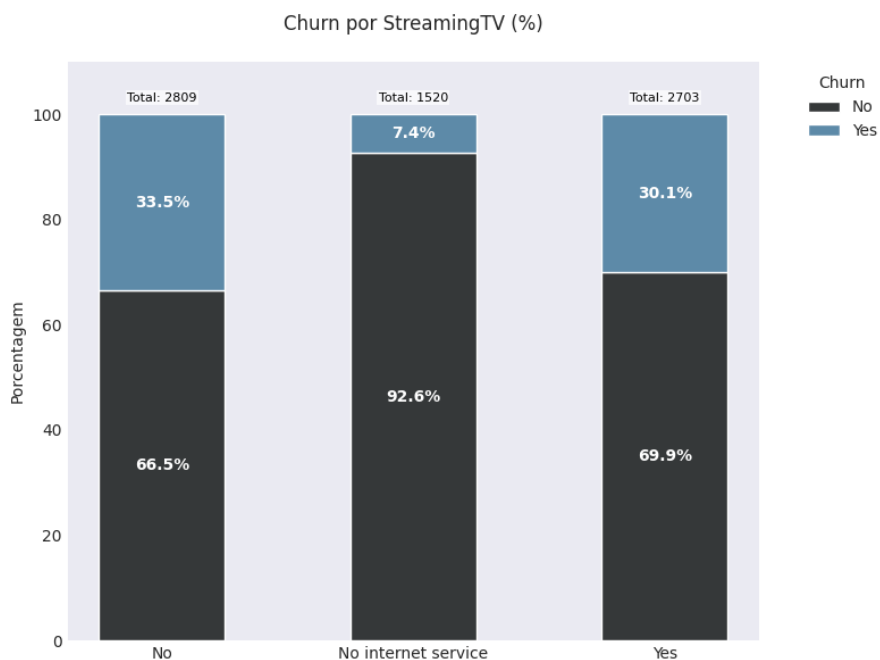


Figura 15: Churn por StreamingTV

O serviço de streaming de TV por assinatura apresenta proporção equivalente de cancelamento entre aqueles que possuem e não possuem o serviço adicional no seu plano, cerca de 30% a 33% de taxa de cancelamento, respectivamente.

5.5 Churn vs Streaming de Filmes

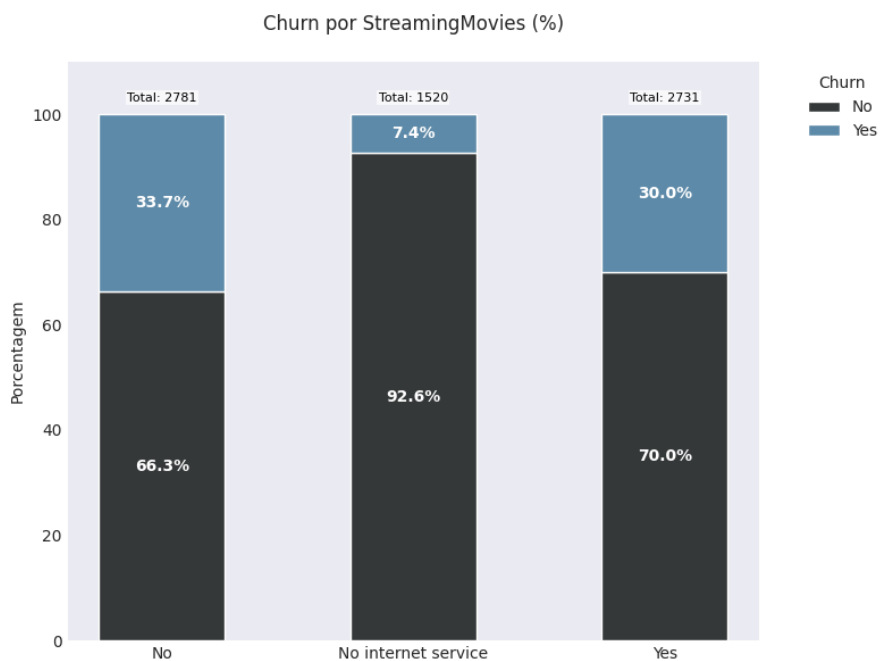


Figura 16: Churn por StreamingMovies

O comportamento é semelhante para aqueles que adicionaram serviço de filmes no plano: entre 30% a 33.7% de taxa de cancelamento, respectivamente, para quem possui e quem não possui o serviço.

6 Grupo 3 - Churn vs (OnlineSecurity | OnlineBackup | DeviceProtection | TechSupport)

6.1 Churn vs Serviço de Segurança Online (OnlineSecurity)

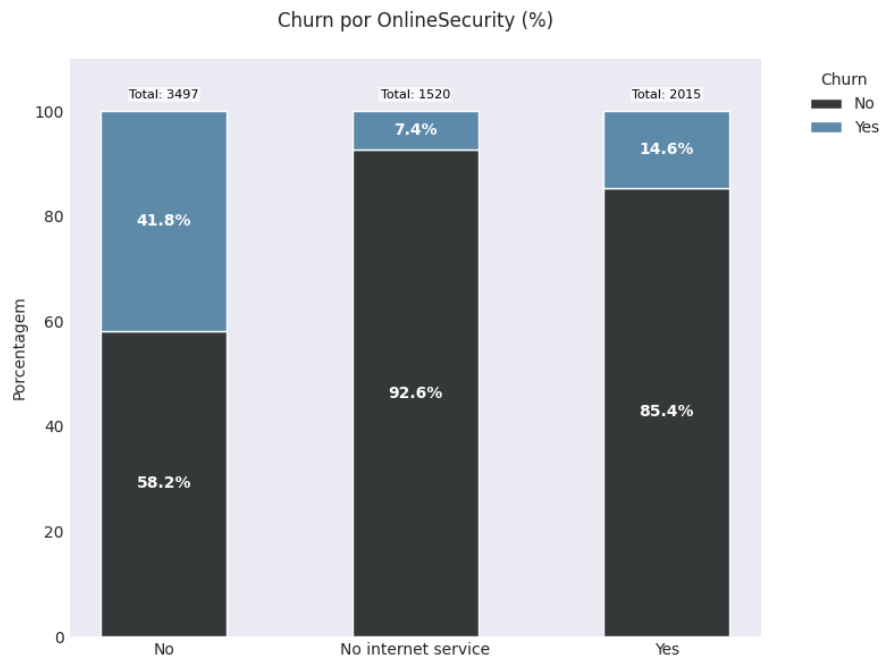


Figura 17: Churn por OnlineSecurity

Clientes que não possuem pacote de segurança online, como VPN e gerenciador de senhas, têm um número considerável de cancelamentos em comparação com os clientes que possuem o serviço. Oferecer diferentes pacotes de proteção na web pode ser uma forma de mitigar o churn.

6.2 Churn vs Serviço de Backup (OnlineBackup)

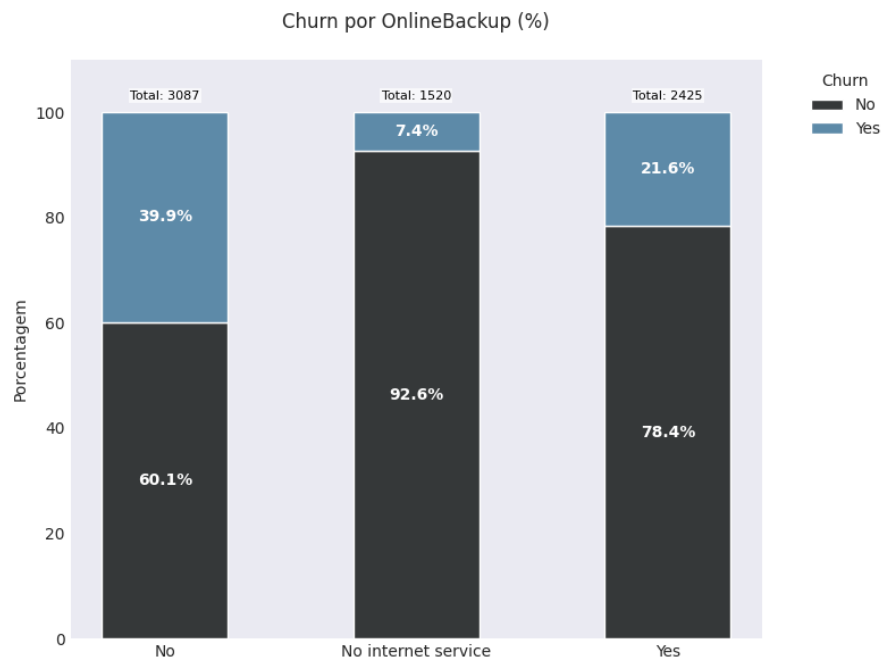


Figura 18: Churn por OnlineBackup

Semelhante ao caso anterior, o número de churn mais que dobra (de 523 para 1233) nos casos de clientes que não possuem serviço de backup de proteção de dados.

6.3 Churn vs Serviço de Proteção para Dispositivos (DeviceProtection)

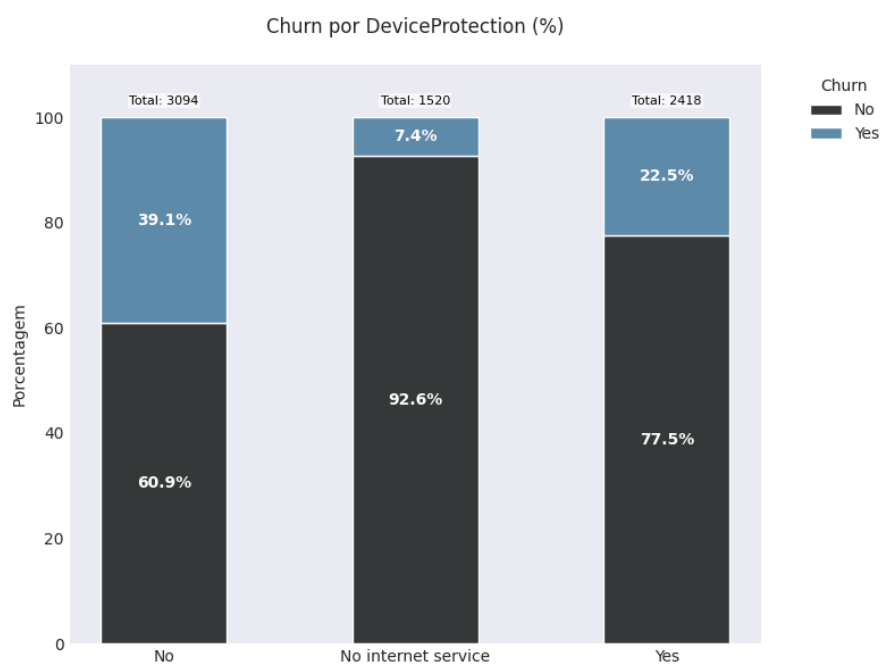


Figura 19: Churn por DeviceProtection

Mais uma vez, podemos observar o dobro de casos de cancelamento entre clientes que não possuem proteção adicional em relação aos que possuem este serviço no pacote.

6.4 Churn vs Suporte Técnico (TechSupport)

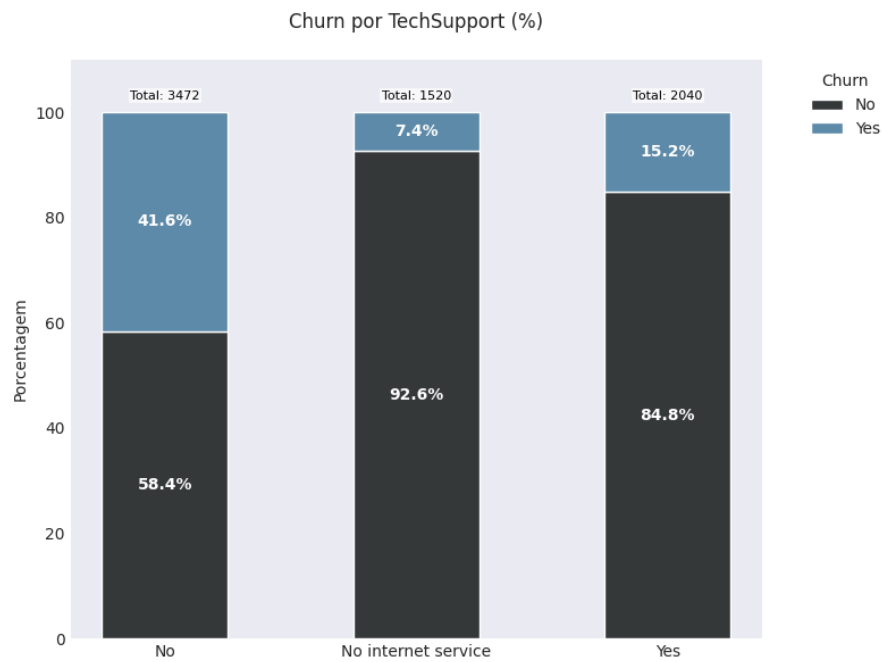


Figura 20: Churn por TechSupport

Entre os clientes que não possuem suporte técnico adequado, há um crescimento expressivo no número de churn: de 310 para 1446, resultando em um aumento de 366.45% nos casos de cancelamento.

7 Grupo 4 - Churn vs (Contract | PaperlessBilling | PaymentMethod)

7.1 Churn vs Método de Pagamento

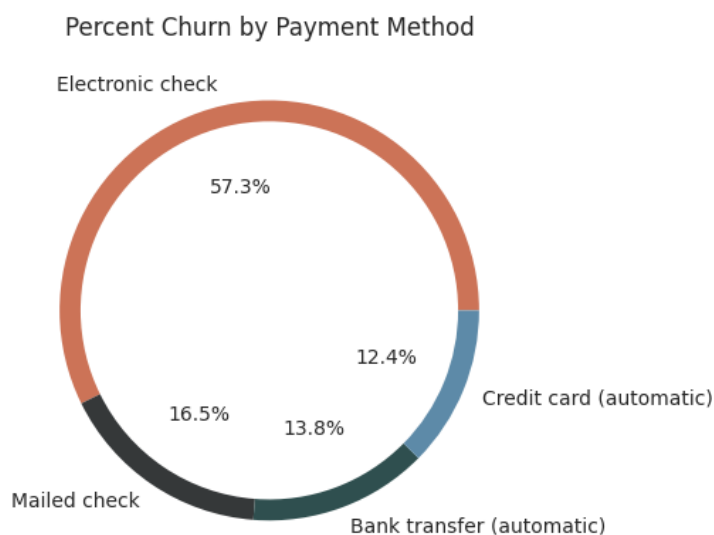


Figura 21: Churn por método de pagamento

Majoritariamente, os clientes que optam pelo método de pagamento com cheque eletrônico apresentam a maior taxa de cancelamento (57.3%), contrastando com o débito automático no cartão de crédito (12.4%). Os métodos automáticos representam maior fidelização do cliente e, conseqüentemente, maior tempo de contrato e menor taxa de cancelamento.

7.2 Churn por Tipo de Contrato

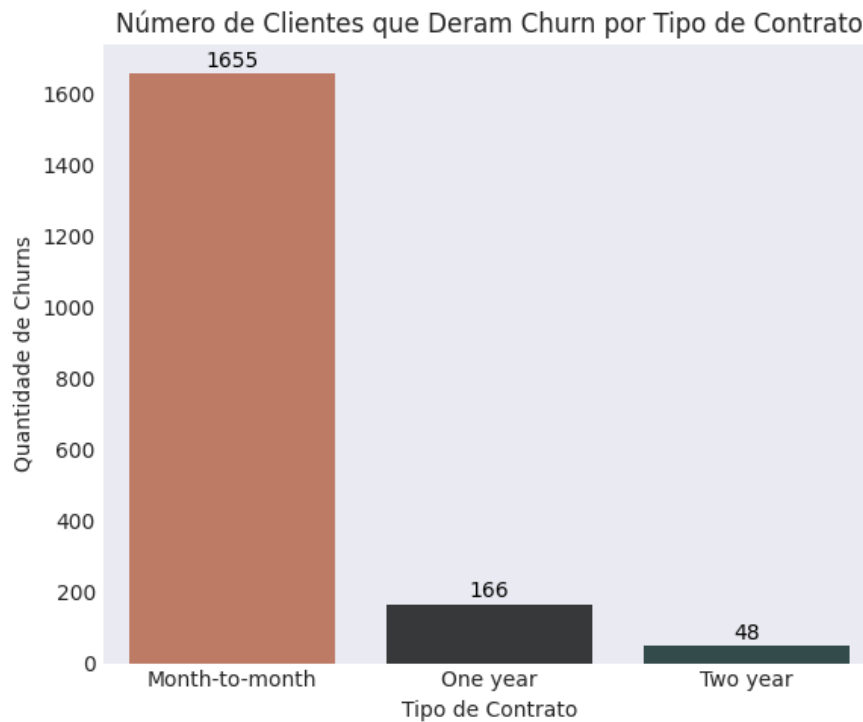


Figura 22: Churn por tipo de contrato

Como visto anteriormente por meio de diferentes análises, o churn é significativamente maior entre clientes com pouco tempo de contrato.

8 Perfis de Risco

A partir da análise exploratória realizada, foi possível identificar diferentes segmentos de clientes com níveis distintos de risco de cancelamento. Estes perfis foram construídos com base nas variáveis mais impactantes sobre o churn, como tempo de contrato, forma de pagamento, serviços adicionais e características demográficas.

Perfil 1: Altíssimo Risco de Churn

- **Contrato:** Mensal
- **Forma de Pagamento:** Cheque Eletrônico
- **Tempo de Contrato:** Inferior a 6 meses
- **Serviços Adicionais:** Sem serviços de segurança online, backup ou suporte técnico
- **Cobrança Total:** Inferior a \$1000
- **Descrição:** Clientes recém-adquiridos, com pouca fidelização e baixa percepção de valor agregado. Estes clientes utilizam métodos de pagamento mais voláteis e não possuem vínculo com serviços adicionais, o que os torna altamente suscetíveis ao churn.

Perfil 2: Risco Moderado de Churn

- **Contrato:** Mensal
- **Forma de Pagamento:** Diversos métodos (inclusive automáticos)
- **Serviços:** Internet por fibra óptica, mas com ausência de pacotes adicionais
- **Demografia:** Clientes sem parceiro(a) e sem dependentes
- **Descrição:** Este perfil reúne clientes que valorizam velocidade, mas não se conectam com serviços adicionais que poderiam aumentar a fidelização. O risco é acentuado para clientes solteiros e sem responsabilidades familiares.

Perfil 3: Risco Potencial

- **Contrato:** Mensal ou anual
- **Serviços:** Com serviços de streaming, porém sem proteção de dispositivos ou suporte técnico
- **Descrição:** Clientes parcialmente engajados com o serviço, porém desprotegidos em termos de suporte e segurança. Podem migrar para concorrentes que ofereçam pacotes mais completos.

Perfil 4: Baixo Risco de Churn

- **Contrato:** Anual ou Bianual
- **Forma de Pagamento:** Débito Automático ou Cartão de Crédito
- **Tempo de Contrato:** Superior a 24 meses
- **Serviços:** Pacotes completos (segurança online, backup, suporte técnico, serviços de streaming)
- **Demografia:** Clientes com parceiro(a) e dependentes
- **Descrição:** Estes clientes apresentam maior tempo de relacionamento com a empresa, múltiplos serviços contratados e métodos de pagamento automáticos, características que demonstram alta fidelização e baixo risco de churn.

Recomendações Estratégicas

- Focar ações de retenção nos clientes do **Perfil 1 e Perfil 2**, oferecendo:
 - Benefícios exclusivos para contratos de maior duração.
 - Pacotes de segurança online e suporte técnico com descontos no início do contrato.
 - Campanhas educacionais para demonstrar o valor agregado dos serviços adicionais.

- Oferecer migração facilitada de métodos de pagamento voláteis para métodos automáticos.
 - Monitorar clientes recém-adquiridos e criar planos de onboarding mais eficazes para melhorar a experiência inicial e reduzir o churn precoce.
-