

4 Análise dos resultados

4.1. Construção do modelo de regressão logística

Após o entendimento do negócio e do objetivo, estruturação dos dados da empresa e definição das informações existentes que possam ajudar para uma melhor caracterização do *churn*, iniciou-se, então, a aplicação do Modelo de Regressão Logística binária no SPSS 13.0.

Conforme já visto anteriormente, a amostra foi dividida em duas partes similares, sendo uma para treinamento e outra para validação (*holdout*), cada uma com 48.963 casos.

O objetivo é buscar a identificação dos terminais com maior probabilidade de serem cancelados, assim temos a variável alvo FLAG_CHURN e 32 variáveis independentes¹ para a avaliação do modelo.

Tabela de Classificação ^{c,d}								
Observado		Previsto						
		Casos selecionados ^a			<i>Holdout</i> ^b			
		FLAG_CHURN		Porcentagem	FLAG_CHURN		Porcentagem	
0	1	Correta	0	1	Correta			
Passo 0	FLAG_CHURN	0	0	34.180	0	0	34.180	0
		1	0	14.783	100	0	14.783	100
Porcentagem Geral					30,2			30,2

a. Casos selecionados HOLD EQ 0

b. Casos não selecionados HOLD NE 0

c. Constante está inclusa no modelo

d. O valor de corte é 0,300

Tabela 13 – Tabela de Classificação – proporção de *churn*.

Como 69,8% da base são de terminais ativos, se arbitrariamente todos os terminais fossem considerados *churners*, a taxa de acerto seria de aproximadamente 30%. O modelo de regressão logística que irá estimar o risco de cancelamento deve ser mais assertivo na classificação dos clientes, do que

¹ Como a utilização das variáveis numéricas, normalizadas pelo *zscore*, dificultaria a interpretação dos coeficientes, estas variáveis (originais) foram transformadas em bandas para aplicação do modelo (conforme quadros 41 a 49 do anexo).

este percentual. Assim, o ponto de corte que parece ser mais adequado deve ser de 70% para terminais ativos e 30% para retirados.

Como foi pontuado no capítulo 3, quando não se conhece a proporção, costuma-se usar o ponto de corte 50%, que define probabilidades iguais para os dois grupos.

Testes de Significância dos Coeficientes do Modelo

		χ^2	df	Sig.
Passo 1	Passo	7.405,859	1	0,000
	Bloco	7.405,859	1	0,000
	Modelo	7.405,859	1	0,000
Passo 2	Passo	3.464,767	5	0,000
	Bloco	10.870,626	6	0,000
	Modelo	10.870,626	6	0,000
Passo 3	Passo	3.459,156	3	0,000
	Bloco	14.329,782	9	0,000
	Modelo	14.329,782	9	0,000
Passo 4	Passo	1.049,488	3	0,000
	Bloco	15.379,270	12	0,000
	Modelo	15.379,270	12	0,000
Passo 5	Passo	345,936	1	0,000
	Bloco	15.725,206	13	0,000
	Modelo	15.725,206	13	0,000
Passo 6	Passo	282,782	15	0,000
	Bloco	16.007,988	28	0,000
	Modelo	16.007,988	28	0,000
Passo 7	Passo	187,310	1	0,000
	Bloco	16.195,298	29	0,000
	Modelo	16.195,298	29	0,000
Passo 8	Passo	96,523	4	0,000
	Bloco	16.291,821	33	0,000
	Modelo	16.291,821	33	0,000
Passo 9	Passo	84,154	6	0,000
	Bloco	16.375,975	39	0,000
	Modelo	16.375,975	39	0,000
Passo 10	Passo	65,325	1	0,000
	Bloco	16.441,301	40	0,000
	Modelo	16.441,301	40	0,000
Passo 11	Passo	8,976	1	0,003
	Bloco	16.450,277	41	0,000
	Modelo	16.450,277	41	0,000
Passo 12	Passo	12,993	1	0,005
	Bloco	16.463,270	44	0,000
	Modelo	16.463,270	44	0,000
Passo 13	Passo	12,465	3	0,006
	Bloco	16.475,734	47	0,000
	Modelo	16.475,734	47	0,000

Tabela 14 – Passos do Método *Stepwise Forward*.

O método utilizado foi o *Stepwise Forward*, onde a cada etapa é incorporada uma variável, que seja mais relevante², de forma a construir o modelo mais estável numericamente, minimizando o número de variáveis, já que quanto mais variáveis incluídas no modelo, mais ele se torna dependente dos dados.

O modelo incorporou 13 passos até chegar ao final (tabela 14). Assim, a primeira variável “quantidade de terminais fixos (QTD_FIXO)” foi contemplada, por ter a maior estatística de Wald. Com esta variável a regressão logística obtém 19,9% de Pseudo R² e *hit ratios* (taxa de sucesso) de 76,8% tanto para o treinamento, quanto para a validação (*holdout*).

Examinando os resultados, entretanto, identificam-se duas razões para a adição de mais variáveis:

1. Ainda há variáveis significantes que podem ser inclusas e possam contribuir para melhorar a adequação do modelo;
2. A taxa de sucesso do *holdout* é boa (76,8%), porém o grupo *churn* apresenta um baixo acerto de 43,4%.

Assim, a segunda variável “segmento (NU_SEGMTO_ATENDTO)”, a terceira “grupo (GRUPO)” e a quarta “tempo do último fixo (TEMPO_ULT_FIXO)”, respectivamente, foram as próximas variáveis inclusas, compondo as quatro juntas 94% do poder explanatório do modelo.

Das 32 variáveis, o modelo final selecionou 13, conforme relacionadas abaixo:

- Quantidade de terminais fixos;
- Segmento (valor do cliente);
- Grupo de produtos (fixo, móvel e banda larga);
- Tempo do último fixo;
- Informações se o cliente já havia tido a intenção de cancelamento anterior e foi retido;
- UF do terminal;
- Disponibilidade de banda larga;
- Idade do cliente;
- Quantidade de meses (1 à 6) de minutos faturados de ligação de fixo para celular;
- Concorrência 1;

² Wald é a estatística utilizada pelo SPSS para incluir ou excluir as variáveis, tanto para o método *forward* como para o *backward*.

- Utilização ou não de internet nos últimos 6 meses;
- Tempo da primeira banda larga;
- Tempo do primeiro móvel.

4.2. Avaliação do ajuste geral do modelo

Nesta etapa será testado o poder explanatório do modelo. Para tanto, será avaliado o pseudo R^2 , as estatísticas de aderência do modelo e o acerto da classificação (*hit ratio*).

A cada passo, tanto o Cox & Snell R^2 quanto Nagelkerke R^2 (conforme tabela 15), aumentam à medida que as variáveis independentes são adicionadas, indicando uma melhora do modelo.

Passo	-2LL	Cox & Snell R^2	Nagelkerke R^2
1	52.572,240 ^a	0,140	0,199
2	49.107,473 ^b	0,199	0,282
3	45.648,317 ^b	0,254	0,359
4	44.598,829 ^b	0,270	0,382
5	44.252,893 ^b	0,275	0,389
6	43.970,111 ^b	0,279	0,395
7	43.782,801 ^b	0,282	0,399
8	43.686,278 ^b	0,283	0,401
9	43.602,124 ^b	0,284	0,403
10	43.536,798 ^b	0,285	0,404
11	43.527,822 ^b	0,285	0,404
12	43.514,829 ^b	0,286	0,404
13	43.502,364 ^b	0,286	0,405

a. estimativa encerrada em 4 iterações, porque os parâmetros estimados mudaram para menos que ,001

b. estimativa encerrada em 5 iterações, porque os parâmetros estimados mudaram para menos de ,001

Tabela 15 – Cox & Snell R^2 , Nagelkerke R^2 e -2LL do modelo.

De forma inversa, a função logaritmo de verossimilhança -2LL da base reduz a cada etapa. Este teste objetiva verificar o poder de influência das variáveis independentes sobre a variável dependente. Menor valor de -2LL indica melhor adequação do modelo.

Passo	χ^2	df	Sig.
1	0,000	0	.
2	3,391	4	0,495
3	212,148	6	0,000
4	198,165	7	0,000
5	247,598	8	0,000
6	276,021	8	0,000
7	295,851	8	0,000
8	283,761	8	0,000
9	274,691	8	0,000
10	253,366	8	0,000
11	259,640	8	0,000
12	255,951	8	0,000
13	247,619	8	0,000

Tabela 16 – Teste estatístico Hosmer e Lemeshow.

O teste Hosmer e Lemeshow (demonstrado na tabela 16) tem a finalidade de avaliar a validade preditiva do modelo de Regressão. Este resultado é apresentado por um valor *chi-quadrado* não significativo, indicando se há diferença estatística significativa entre as classificações observadas e previstas da variável dependente, ou seja, quanto menor a diferença na classificação, melhor o modelo.

Embora este teste indique que há diferenças entre os valores reais e os esperados, o modelo demonstra um bom percentual de acerto de 79,2% no último passo (tabela 17).

Na etapa final do modelo, todas as medidas de ajuste melhoraram. Como pode ser observado na tabela 15, o valor -2LL reduziu de 52.572 para 43.502, e os valores do pseudo R^2 apresentaram um crescimento de 0,199 para 0,405. Embora este valor de R^2 seja ainda distante do ideal, não invalida o modelo.

4.3. Precisão da estimativa

A matriz de classificação representa o nível de precisão da estimativa, resultado da aplicação do modelo. A medida utilizada para alcançar este resultado é o *hit ratio* (taxa de sucesso), que é o percentual de casos corretamente classificados.

O método de Regressão Logística é direcionado para maximizar a probabilidade, o que não quer dizer ser a melhor precisão da estimativa, portanto, pode haver problemas na utilização deste modelo quando o objetivo da análise é a precisão.

Tabela de Classificação ^c								
Observado	Previsto							
	Casos selecionados ^a			Holdout ^b				
	FLAG_CHURN		Porcentagem	FLAG_CHURN		Porcentagem		
	0	1	Correta	0	1	Correta		
Passo 1	FLAG_CHURN	0	31.189	2.991	91,2	31.277	2.903	91,5
		1	8.362	6.421	43,4	8.463	6.320	42,8
Porcentagem Geral					76,8			76,8
Passo 2	FLAG_CHURN	0	31.109	3.071	91,0	31.198	2.982	91,3
		1	7.144	7.639	51,7	7.241	7.542	51,0
Porcentagem Geral					79,1			79,1
Passo 3	FLAG_CHURN	0	26.995	7.185	79,0	27.015	7.165	79,0
		1	3.752	11.031	74,6	3.728	11.055	74,8
Porcentagem Geral					77,7			77,8
Passo 4	FLAG_CHURN	0	27.840	6.340	81,5	27.882	6.298	81,6
		1	4.072	10.711	72,5	4.044	10.739	72,6
Porcentagem Geral					78,7			78,9
Passo 5	FLAG_CHURN	0	27.877	6.303	81,6	27.902	6.278	81,6
		1	4.045	10.738	72,6	3.977	10.806	73,1
Porcentagem Geral					78,9			79,1
Passo 6	FLAG_CHURN	0	27.829	6.351	81,4	27.885	6.295	81,6
		1	4.020	10.763	72,8	4.029	10.754	72,7
Porcentagem Geral					78,8			78,9
Passo 7	FLAG_CHURN	0	27.863	6.317	81,5	27.951	6.229	81,8
		1	4.014	10.769	72,8	4.033	10.750	72,7
Porcentagem Geral					78,9			79,0
Passo 8	FLAG_CHURN	0	27.864	6.316	81,5	27.907	6.273	81,6
		1	3.971	10.812	73,1	4.009	10.774	72,9
Porcentagem Geral					79,0			79,0
Passo 9	FLAG_CHURN	0	27.889	6.291	81,6	27.934	6.246	81,7
		1	3.983	10.800	73,1	3.991	10.792	73,0
Porcentagem Geral					79,0			79,1
Passo 10	FLAG_CHURN	0	27.918	6.262	81,7	27.951	6.229	81,8
		1	3.965	10.818	73,2	3.986	10.797	73,0
Porcentagem Geral					79,1			79,1
Passo 11	FLAG_CHURN	0	27.914	6.266	81,7	27.939	6.241	81,7
		1	3.977	10.806	73,1	3.977	10.806	73,1
Porcentagem Geral					79,1			79,1
Passo 12	FLAG_CHURN	0	27.914	6.266	81,7	27.946	6.234	81,7
		1	3.966	10.817	73,2	3.979	10.804	73,1
Porcentagem Geral					79,1			79,1
Passo 13	FLAG_CHURN	0	27.917	6.263	81,7	27.933	6.247	81,7
		1	3.945	10.838	73,3	3.965	10.818	73,2
Porcentagem Geral					79,2			79,1

a. Casos selecionados HOLD EQ 0

b. Casos não selecionados HOLD NE 0

c. O valor de corte é 0,300

Tabela 17 – Matriz de classificação das amostras de treinamento e de validação.

Após a aplicação das 13 variáveis a regressão chegou a uma taxa de acerto geral de 79,2%, ou seja, um bom nível de acerto. Desse, as taxas de acerto foram 73,3% e 81,7%, para *churn* e ativo, respectivamente.

O modelo inicial, que considerava apenas a constante, apresentava uma taxa de sucesso de 30,2%. Após o décimo terceiro passo obtém-se uma evolução no acerto da previsão de 2,32 vezes.

A partir do passo 11, a melhora na taxa de acerto geral é muito pequena e o pseudo R^2 não se altera. Assim poderíamos descartar do modelo final as variáveis 11, 12 e 13, entretanto, as mesmas foram mantidas para ajudar na definição do perfil dos clientes *churners*.

4.4. Validação do modelo (*Holdout Sample*)

Para validação da eficiência classificatória do modelo, como já citado anteriormente, a amostra foi dividida em duas partes iguais, extraída através de geração de números aleatórios:

1. amostra para treinamento do modelo - 48.963 casos; e
2. amostra para validação do modelo, denominada *holdout sample* – 48.963 casos.

Como observado na tabela 17, as taxas de sucesso da amostra de validação são quase idênticas às taxas de acerto da amostra de treinamento. Assim há indicativos que o modelo de regressão logística possui suporte empírico para as duas amostras – treinamento e validação.

4.5. Interpretação dos resultados

Conforme visto no capítulo 3, o Modelo de Regressão Logística permite estimar a probabilidade de ocorrência de um evento e identificar as variáveis independentes que contribuem efetivamente para a sua predição. O evento de interesse aqui é a previsão de *churn* de clientes de telefonia fixa e foi composto pelas variáveis:

- Quantidade de terminais fixos;
- Segmento (valor do cliente);
- Grupo de produtos (fixo, móvel e banda larga);
- Tempo do último fixo;
- Informações se o cliente já havia tido a intenção de cancelamento anterior e foi retido;
- UF do terminal;

- Disponibilidade de banda larga;
- Idade do cliente;
- Quantidade de meses (1 a 6) de minutos faturados de ligação de fixo para celular;
- Concorrência 1;
- Utilização ou não de internet nos últimos 6 meses;
- Tempo da primeira banda larga;
- Tempo do primeiro móvel.

O evento *churn* pode ser motivado por diversos fatores. Alguns deles podem ser descritos por meio de variáveis e utilizados para estabelecer uma relação entre a ocorrência ou não dos fatores e o evento. As variáveis acima foram selecionadas pelo modelo por apresentarem influência sobre a diferenciação entre os grupos ativos e *churners*.

Após a geração do Modelo de Regressão Logística, o mesmo atribui uma probabilidade de cancelamento a cada indivíduo da base em estudo, dividindo esta base nos dois grupos (cancelados e ativos). Este processo é representado por um *score* de probabilidade de *churn* para cada um dos clientes da base de dados.

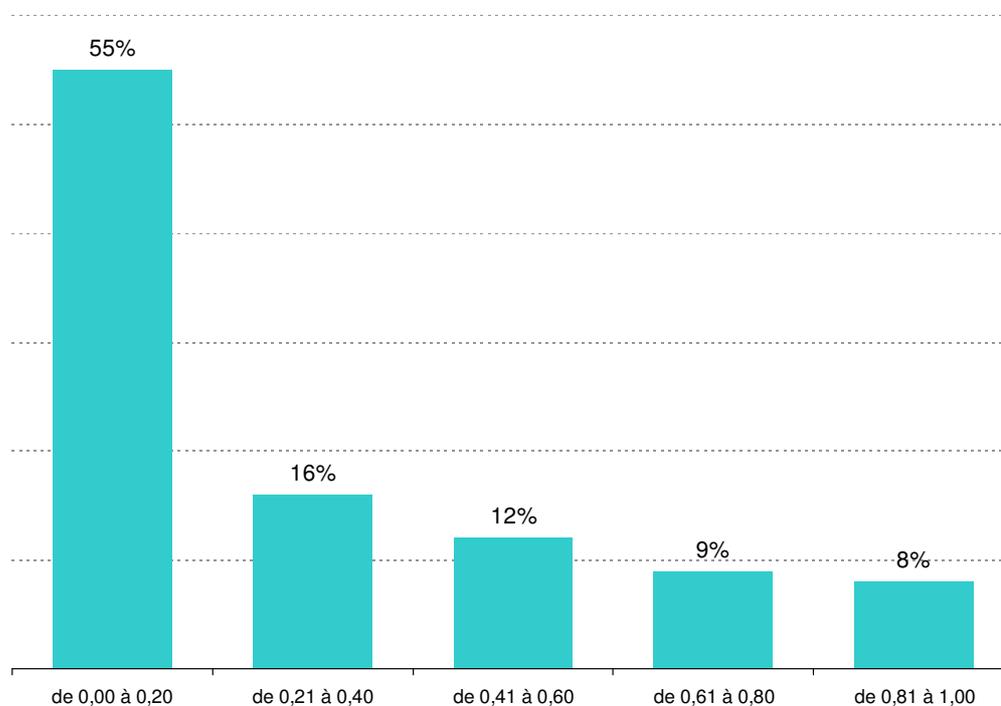


Gráfico 3 – Distribuição da base por *score* de risco de cancelamento.

Com a distribuição da base pelo *score* gerado pelo modelo, é possível observar que a maior parte dos clientes está concentrada em valores mais baixos, ou seja, com menor probabilidade de *churn*. Somente 8% da base de clientes ativos têm mais de 80% de chance de solicitarem cancelamento.

Com esta informação é possível verificar a viabilidade da organização aplicar ações de incentivos preventivos, antecipando-se e tornando mais eficiente e econômica a atividade de reter estes clientes.

Pelos resultados gerados também é possível identificar o perfil dos clientes que cancelam, o que será descrito a seguir, de acordo com o que foi executado e as variáveis que entraram no modelo.

Variável independente 1: **Quantidade de terminais fixos na residência do cliente.**

Dividimos a base em dois grupos: clientes com residências com um fixo e clientes com residência com mais de um fixo.

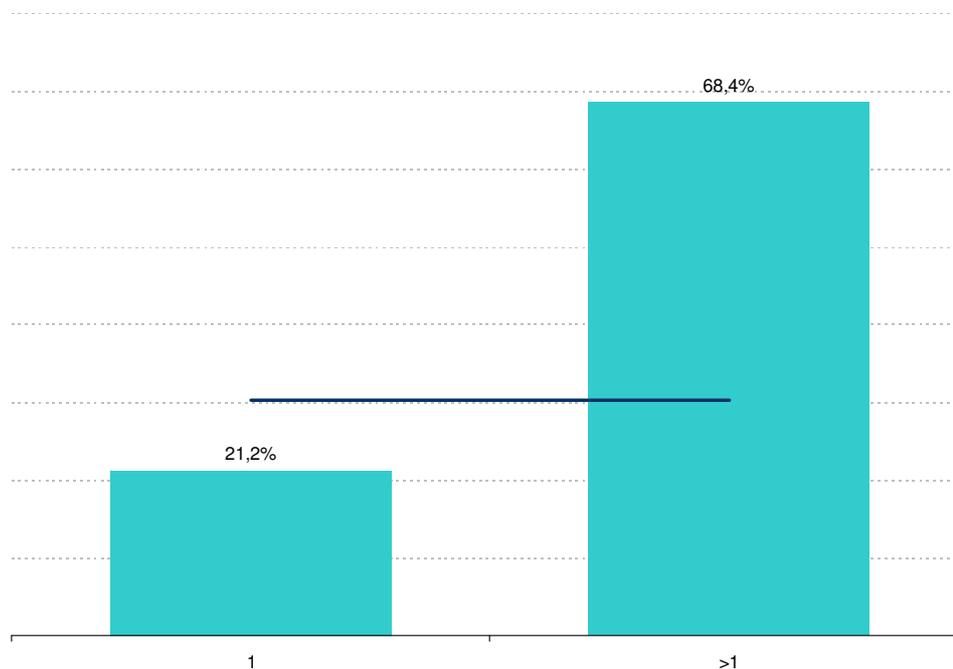


Gráfico 4 – Distribuição da amostra de *churn* por quantidade de terminais fixos na residência do cliente.

Possuir um terminal apresenta menos chance de cancelamento que ter mais de um. O cliente que possui somente um terminal em sua residência apresenta 9,5 vezes menos chance de ser cancelado em relação ao que possui mais de um terminal.

QUANTIDADE DE FIXOS	B	S.E.	Wald	df	Sig.	Exp(B)
1 terminal	(2,25)	0,03	4.572	1	-	0,105

Tabela 18 – Coeficiente da variável quantidade de fixos.

A entrada de novas tecnologias, como por exemplo a banda larga, que permite a utilização de internet sem ocupar a linha telefônica, influencia bastante esta variável.

Variável independente 2: **Segmento (valor) do cliente.**

A base de clientes apresenta clientes diamante, ouro, prata, bronze, novos clientes e sem segmento³.

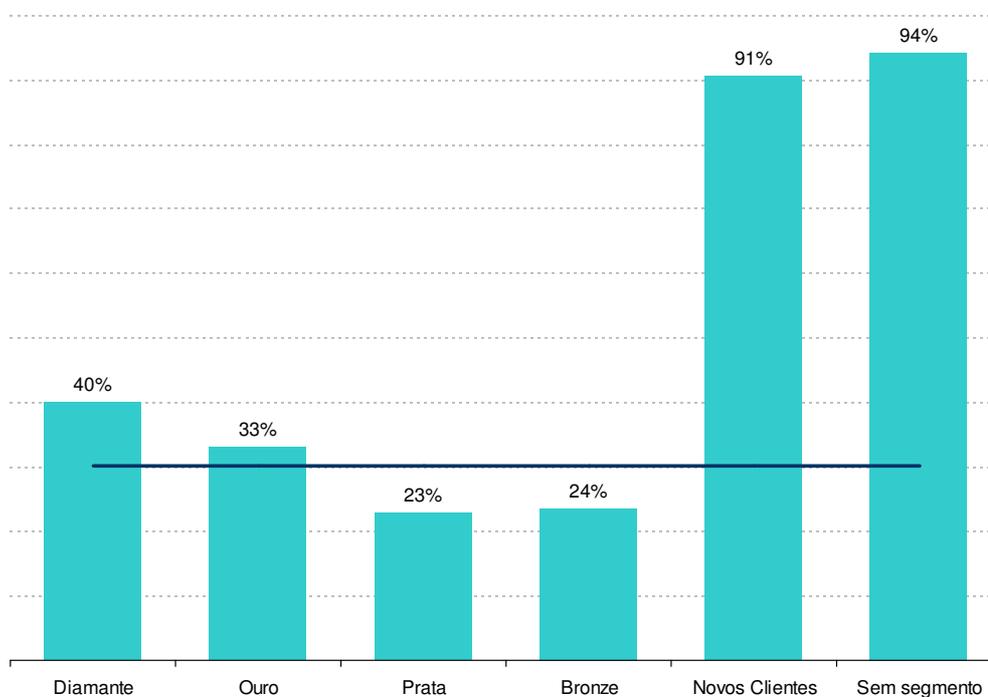


Gráfico 5 – Distribuição da amostra de *churn* por segmento do cliente.

³ Sem segmento - clientes que ainda não foram classificados ou, que por algum motivo, perderam a classificação. Na maioria são clientes novos.

O coeficiente desta variável indica que os clientes sem segmento e novos são os que apresentam maior probabilidade de *churn*. Quando comparados os segmentos diamante, ouro, prata, bronze e novos clientes com o sem segmento, encontra-se, respectivamente, 63,2; 66,12; 72,0; 57,6 e 2,42 vezes menos chance de cancelamento.

Destaca-se também o fato do cliente diamante apresentar probabilidade de *churn* maior do que os demais segmentos, ficando menor apenas do que novos clientes e sem segmento, mesmo sendo, em teoria, os clientes que a empresa foca mais esforços em melhorar a satisfação. Conforme apresentado no referencial teórico por Reichheld (1996a), clientes satisfeitos podem sim solicitar cancelamento. Devido ao seu potencial de consumo, o número de assédio por concorrentes a clientes com este perfil é alto, provocando um maior estímulo.

Clientes sempre avaliarão qual oferta proporciona maior valor, buscando maximizar a percepção, dentro dos limites estabelecidos por custo, procura, conhecimento e receita. Assim, a satisfação pode evitar em parte a perda de clientes, mas não cria vantagem competitiva, por si só (KOTLER, 2000).

SEGMENTO	B	S.E.	Wald	df	Sig.	Exp(B)
Sem segmento			1.212	5	0,00	
Diamante	(4,15)	0,13	1.002	1	0,00	0,016
Ouro	(4,19)	0,13	1.042	1	0,00	0,015
Prata	(4,28)	0,13	1.107	1	0,00	0,014
Bronze	(4,05)	0,13	997	1	0,00	0,017
Novos Clientes	(0,89)	0,35	6	1	0,01	0,413

Tabela 19 – Coeficientes da variável segmento.

Variável independente 3: **Grupo de produtos.**

São quatro os grupos de produtos avaliados: 1. somente fixo, 2. fixo + móvel, 3. fixo + banda larga e 4. fixo + móvel + banda larga.

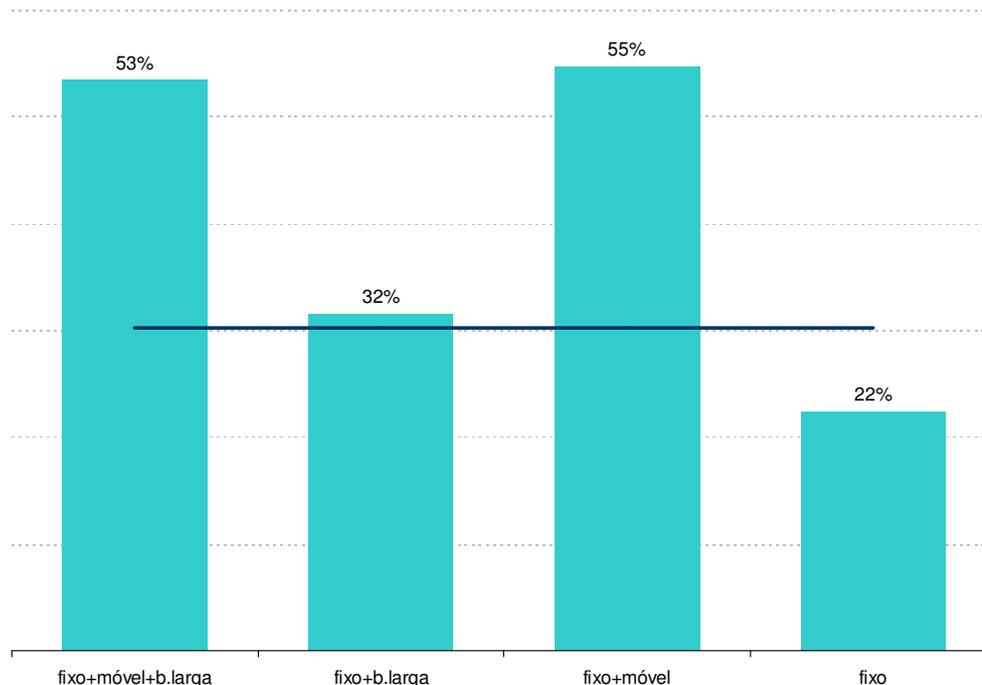


Gráfico 6 – Distribuição da amostra de churn por grupo de produtos/serviços.

Clientes que possuem outros produtos além do fixo têm maior propensão ao *churn*. Isto demonstra a facilidade de troca por outra tecnologia ou concorrente, principalmente em ofertas *bundle*⁴. Exceto pelo produto de fixo + banda larga, que não se pode afirmar que existam diferenças significativas⁵, ter fixo + móvel ou fixo + móvel + banda larga demonstram chances, respectivas de cancelamento, 9,35 e 7,61 vezes maior do que ter somente o fixo.

GRUPO	B	S.E.	Wald	df	Sig.	Exp(B)
Fixo			49	3	0,00	
Fixo + Móvel + Banda Larga	2,03	1,00	4	1	0,04	7,610
Fixo + Banda Larga	0,21	0,73	0	1	0,77	1,237
Fixo + Móvel	2,24	0,68	11	1	0,00	9,353

Tabela 20 – Coeficientes da variável grupo de produtos/serviços.

⁴ Oferta *bundle* - Oferta em pacote: cada vez mais utilizadas pelas empresas para gerar valor para o cliente, aumentando a competitividade. Exemplos: pacotes com telefonia fixa + móvel + banda larga ou telefonia fixa + banda larga + TV, entre outras.

⁵ O grau de significância entre fixo e fixo + banda larga foi alto, desta forma, não há evidências suficientes para afirmar que existam diferenças entre os dois grupos.

Variável independente 4: **Tempo do último fixo.**

A base foi dividida em quatro períodos: 1. até dois anos (1 a 730 dias); 2. de dois a cinco anos (731 a 1.825 dias); 3. de cinco a dez anos (1.826 a 3.650 dias) e 4. mais de 10 anos (mais de 3.650 dias). É importante ressaltar, que para os casos onde o cliente possui apenas um terminal, o tempo constante na variável é o do primeiro fixo.

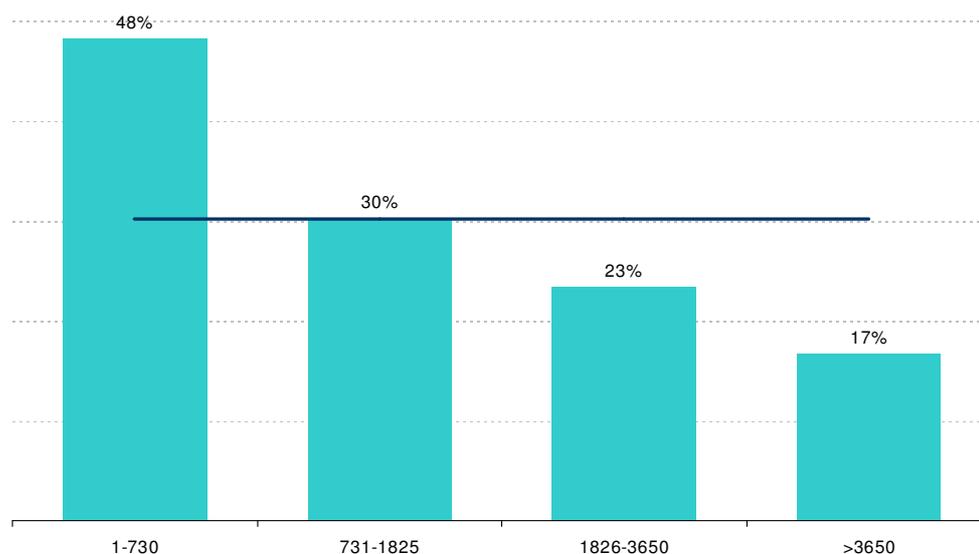


Gráfico 7 – Distribuição da amostra de *churn* por tempo do último fixo.

Esta variável apresentou uma alta correlação com tempo do primeiro fixo, e a mesma foi preferida por ter maior influencia sobre o *churn*. Quanto mais novo o terminal do cliente maior a probabilidade de cancelamento. A relação é de 2,8; 2,1 e 1,6 vezes maior, para de 1 a 2 anos, de 2 a 5 anos e de 5 a 10 anos, respectivamente, do que o que tem mais de 10 anos de base.

TEMPO DO ÚLTIMO FIXO	B	S.E.	Wald	df	Sig.	Exp(B)
> 3.650 dias			712	3	0,00	
1 - 730 dias	1,05	0,04	650	1	0,00	2,853
731 - 1.825 dias	0,73	0,04	318	1	0,00	2,077
1.826 - 3.650 dias	0,44	0,04	134	1	0,00	1,559

Tabela 21 – Coeficientes da variável tempo do último fixo.

Este resultado reflete bem a escala do ciclo de vida do cliente, onde para clientes habituais (mais antigos), uma falha não o leva automaticamente ao abandono, o mesmo avaliará outros fatores antes de tomar sua decisão. Isto não é verdade, quando se trata de novos relacionamentos, momento onde o cliente avalia o desempenho da empresa e/ou serviços, apresentando baixa retenção potencial. Neste momento o consumidor precisa receber informações e perceber boa relação custo versus benefício para continuar comprando (BLATTBERG, GETZ e THOMAS, 2001).

Variável independente 5: **O cliente já havia tido a intenção de cancelamento anterior e foi retido.**

Esta variável indica que o cliente já havia tido não só a intenção de cancelar, como também chegou a solicitar o cancelamento nos últimos 6 meses; entretanto, recebeu uma oferta de retenção e concordou em permanecer na empresa.

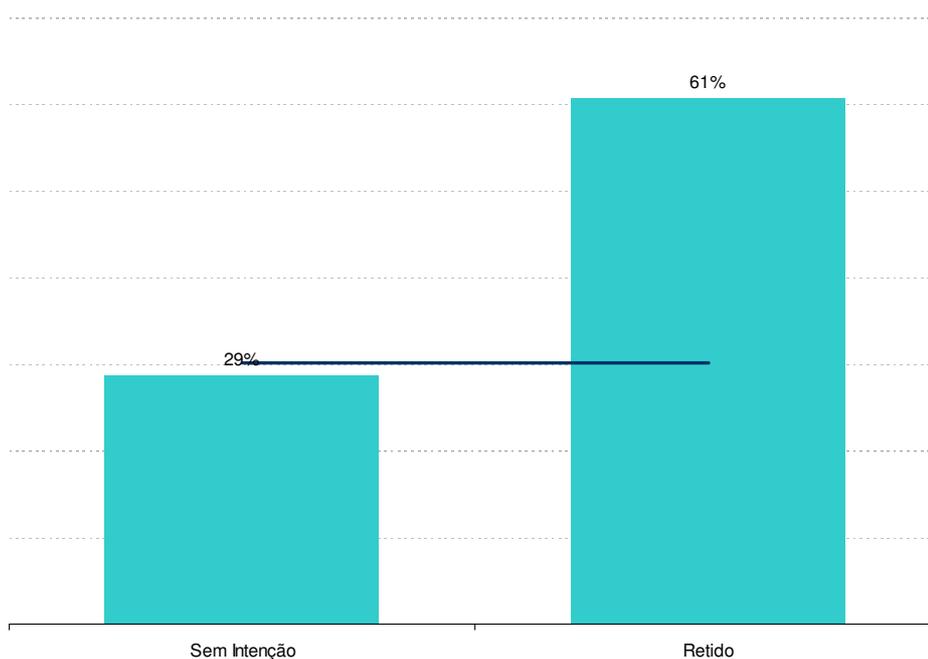


Gráfico 8 – Distribuição da amostra de *churn* por intenção anterior de cancelamento (nos últimos 6 meses).

Se o cliente já apresentou intenção anterior de cancelar, indica que o mesmo já não apresentava laços tão fortes com a empresa. Mesmo com uma retenção, que está indicando não ser eficiente para fortalecer estes laços, a mesma pode estar prolongando o tempo de sobrevivência, mas não resolvendo o real problema.

INTENÇÃO CANCELAMENTO	B	S.E.	Wald	df	Sig.	Exp(B)
Sem intenção e retenção anterior	(0,92)	0,05	295	1	0,00	0,398

Tabela 22 – Coeficiente da variável intenção anterior de cancelamento (nos últimos 6 meses).

Clientes que não manifestaram intenção de cancelamento anterior com retenção têm 2,5 vezes menor probabilidade de *churn* do que os que já o fizeram.

Variável independente 6: **UF**.

Estados onde os terminais estão instalados.

Esta foi a variável que apresentou maior número de coeficientes sem diferenças significativas.

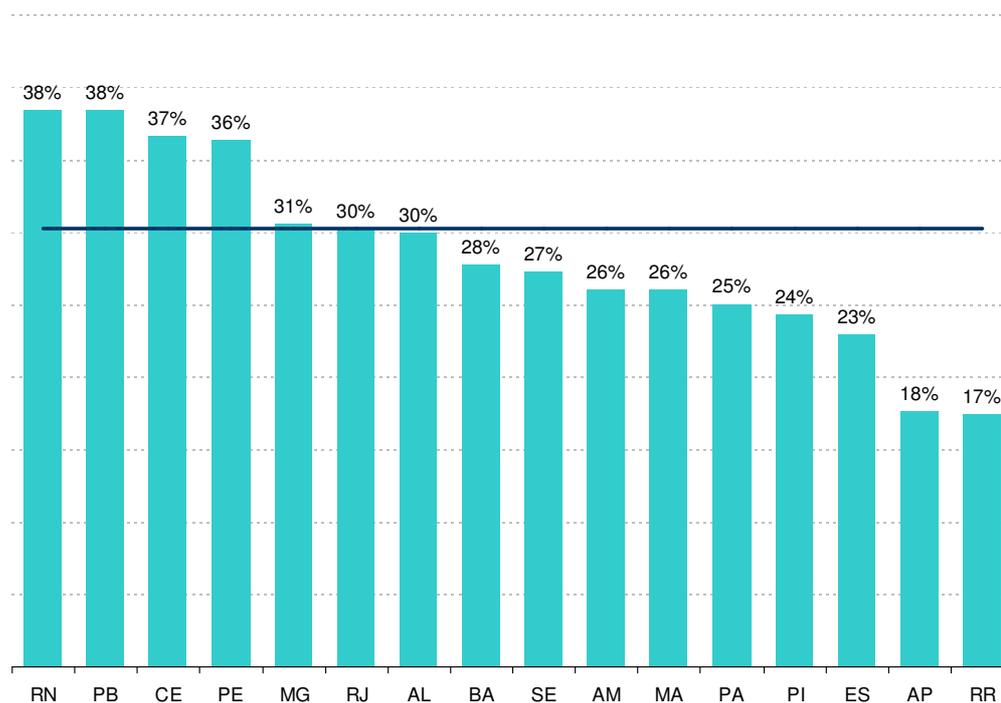


Gráfico 9 – Distribuição da amostra de *churn* por UF.

Espírito Santo, Rio de Janeiro, Sergipe e Bahia, com respectivos 2,4; 2,0; 1,9 e 1,9, apresentam menor chance de cancelamento, quando comparados com Rio Grande do Norte.

ESTADO	B	S.E.	Wald	df	Sig.	Exp(B)
RN			290	15	0,00	
BA	(0,62)	0,08	56	1	0,00	0,539
ES	(0,88)	0,10	78	1	0,00	0,417
RJ	(0,68)	0,08	79	1	0,00	0,505
PA	(0,58)	0,10	33	1	0,00	0,561
SE	(0,63)	0,14	20	1	0,00	0,534
MG	(0,37)	0,08	24	1	0,00	0,688
PI	(0,57)	0,13	21	1	0,00	0,564
AM	(0,52)	0,11	24	1	0,00	0,593
PE	(0,17)	0,09	4	1	0,05	0,843
RR	(0,48)	0,23	4	1	0,03	0,619
AP	(0,54)	0,22	6	1	0,01	0,583
CE	(0,17)	0,09	4	1	0,06	0,844
AL	(0,27)	0,12	5	1	0,03	0,764
PB	0,01	0,11	0	1	0,92	1,010
MA	(0,43)	0,10	17	1	0,00	0,653

Tabela 23 – Coeficientes da variável UF.

Variável independente 7: **Disponibilidade de banda larga.**

Esta variável indica se, na região onde está instalado o terminal, tem ou não disponibilidade para banda larga da empresa.

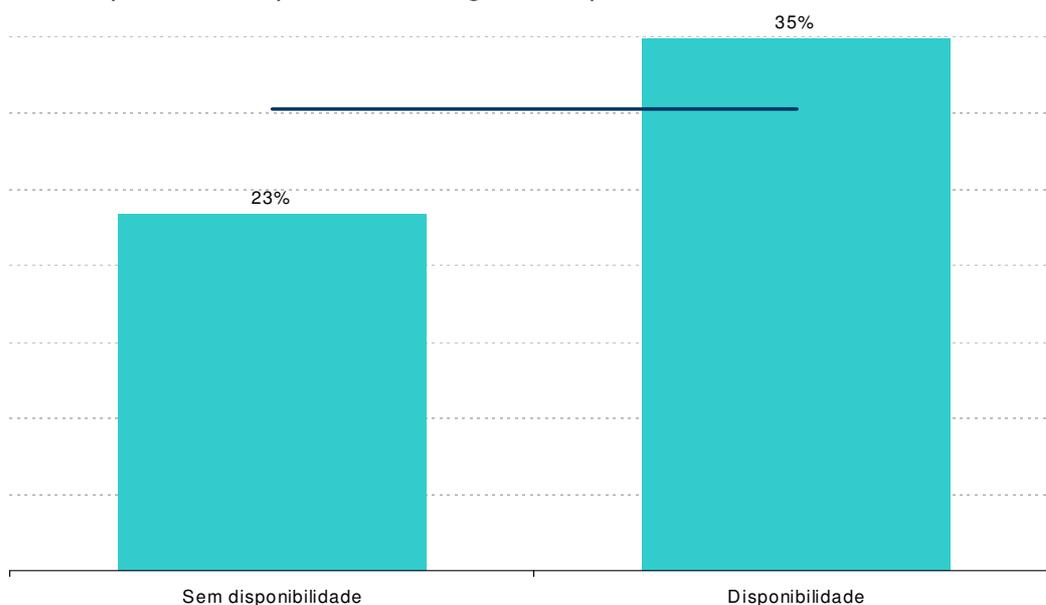


Gráfico 10 – Distribuição da amostra de *churn* por disponibilidade de banda larga.

Onde não há disponibilidade para banda larga apresenta menor probabilidade de cancelamento (1,4 vezes menor). Geralmente, há disponibilidade nas regiões de maior concorrência.

BANDA LARGA	B	S.E.	Wald	df	Sig.	Exp(B)
Sem Disponibilidade de banda larga	(0,34)	0,03	146	1	0,00	0,712

Tabela 24 – Coeficiente da variável disponibilidade de banda larga da empresa.

Variável independente 8: **Idade do cliente.**

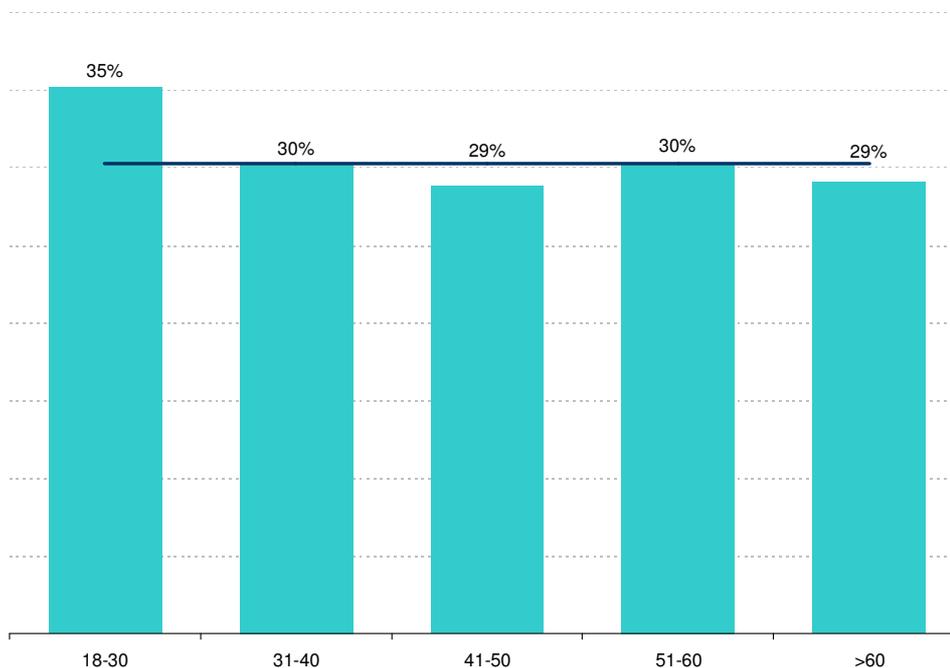


Gráfico 11 – Distribuição da amostra de churn por idade do cliente.

A variável idade do cliente foi dividida em cinco níveis: de 18 a 30 anos, de 31 a 40 anos, de 41 a 50 anos, de 51 a 60 anos e maior que 60 anos.

Os desejos e habilidades dos consumidores mudam com a idade, como também são determinados pelo estágio no ciclo de vida (KOTLER, 2000 e BOONE & KURTZ, 2006). O público mais jovem pode sofrer mais influência por novas tecnologias, substituindo mais facilmente a telefonia fixa por móvel ou outros meios de comunicação.

Dentro desta linha os resultados encontrados são convergentes: quanto maior a idade do cliente, menor a probabilidade de cancelamento. Clientes de 18 a 30 anos apresentam 1,47 vezes mais chance de *churn*, quando comparados aos que possuem mais de 60 anos de idade.

IDADE DO CLIENTE	B	S.E.	Wald	df	Sig.	Exp(B)
> 60 anos			92	4	0,00	
18 - 30 anos	0,38	0,05	71	1	0,00	1,467
31 - 40 anos	0,32	0,04	67	1	0,00	1,379
41 - 50 anos	0,19	0,04	28	1	0,00	1,215
51 - 60 anos	0,17	0,04	20	1	0,00	1,181

Tabela 25 – Coeficientes da variável idade do cliente.

Variável independente 9: **Quantidade de meses de faturamento de VC1⁶**.

Aqui foi utilizado o histórico dos últimos 6 meses de faturamento de utilização do serviço de ligação para celular (fixo – móvel).



Gráfico 12 – Distribuição da amostra de *churn* por quantidade de meses de faturamento de ligação para celular.

⁶ VC1 – ligação de fixo para móvel.

Para a utilização de ligação para celular, apresenta maior probabilidade de cancelamento quem utiliza com menor frequência, de 1 a 4 meses, sendo em torno de 1,3 vezes maior do que quem utilizou pelos 6 meses avaliados.

QUANTIDADE DE MESES VC1	B	S.E.	Wald	df	Sig.	Exp(B)
6 meses de faturamento de vc1			84	6	0,00	
Sem faturamento de vc1	0,16	0,03	23	1	0,00	1,169
1 mês de faturamento de vc1	0,29	0,06	23	1	0,00	1,343
2 meses de faturamento de vc1	0,28	0,06	23	1	0,00	1,329
3 meses de faturamento de vc1	0,25	0,06	20	1	0,00	1,281
4 meses de faturamento de vc1	0,34	0,05	44	1	0,00	1,406
5 meses de faturamento de vc1	0,24	0,04	33	1	0,00	1,271

Tabela 26 – Coeficientes da variável quantidade de meses de faturamento de ligação para celular.

Variável independente 10: **Concorrência 1.**

Esta variável representa a atuação mais forte do principal concorrente para a empresa.

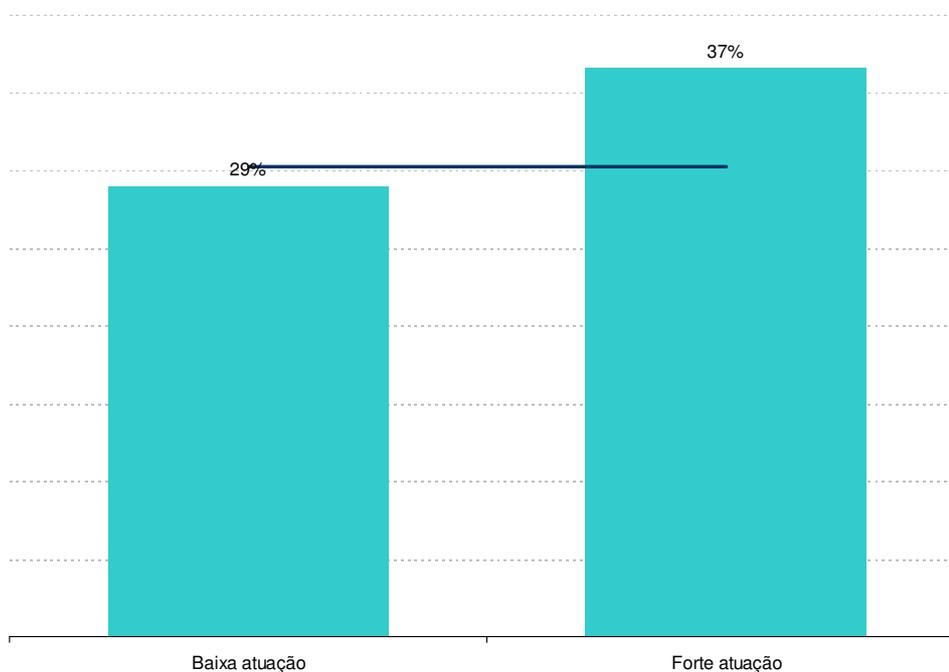


Gráfico 13 – Distribuição da amostra de churn por atuação do concorrente 1.

A presença de terminais em áreas de forte atuação do principal concorrente apresenta 1,3 vezes mais probabilidade de cancelamento do que outras com menor atuação.

CONCORRENTE 1	B	S.E.	Wald	df	Sig.	Exp(B)
Baixa atuação	(0,29)	0,04	62	1	0,00	0,751

Tabela 27 – Coeficiente da variável atuação do concorrente 1.

A portabilidade trouxe novo ânimo para concorrência no setor de telefonia fixa, aumentando a percepção no consumidor de baixo custo de troca de prestadora de serviço (ASAARI e KARIA, 2000), já que agora o mesmo não perde mais o “seu número”, numa eventual mudança (TELECO, 2009).

Este cenário beneficia principalmente aos novos entrantes, como a TIM e NET na telefonia fixa e a entrada da GVT na região I (TELECO, 2009).

Variável independente 11: **Utilização de internet discada nos últimos 6 meses.**

Esta variável indica que o cliente utilizou internet discada nos últimos 6 meses.

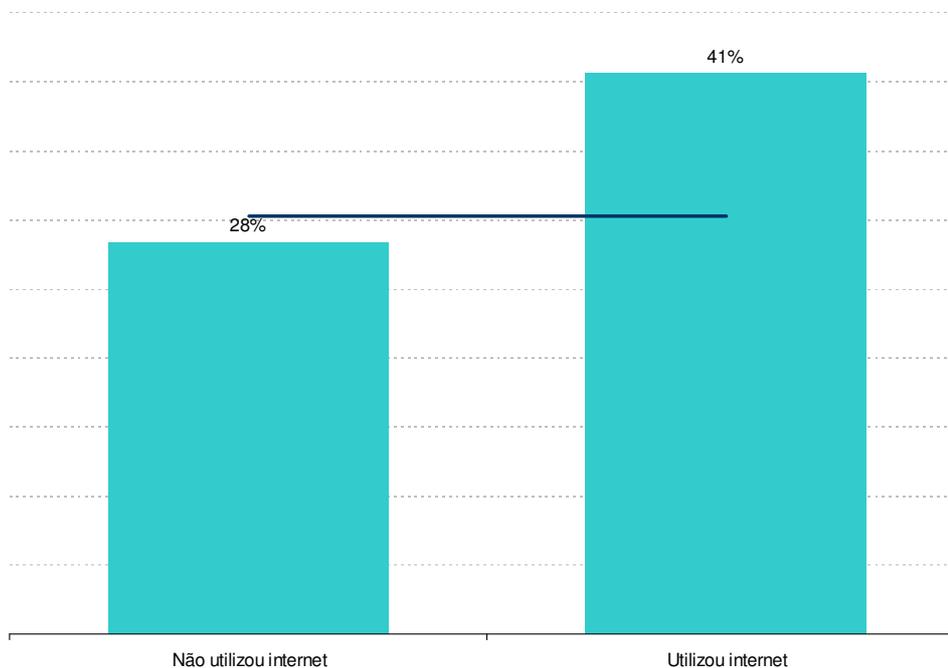


Gráfico 14 – Distribuição da amostra de churn por uso ou não de internet discada.

A variável utilização de internet discada gera uma probabilidade de cancelamento 1,1 vezes maior do que entre aqueles que não a utilizam.

UTILIZAÇÃO DE INTERNET	B	S.E.	Wald	df	Sig.	Exp(B)
Não utilizou nos últimos 6 meses	(0,11)	0,03	10	1	0,00	0,900

Tabela 28 – Coeficiente da variável utilização de internet discada.

Como as variáveis grupo de produtos (variável 3), disponibilidade de banda larga (variável 7) e concorrência 1 (variável 10), esta também sofre influência do aumento de competitividade, principalmente por ofertas convergentes (fixo mais banda larga, por exemplo), proporcionado pelo avanço da tecnologia e redução de preço.

Variável independente 12: Tempo da primeira banda larga.

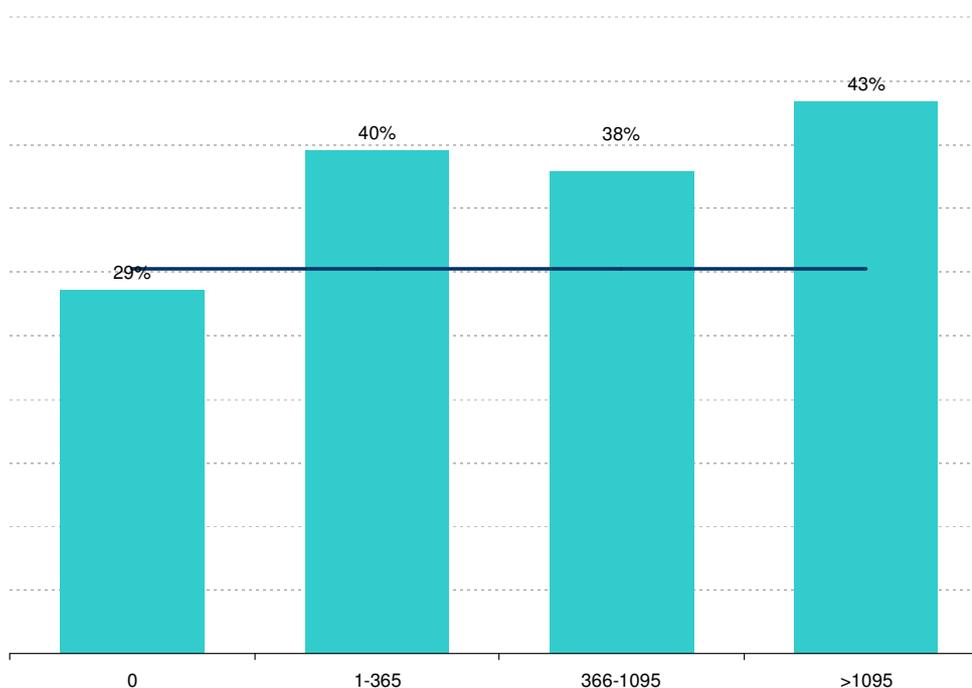


Gráfico 15 – Distribuição da amostra de churn por tempo da primeira banda larga.

Clientes que possuem banda larga com mais de 3 anos (mais de 1.095 dias) apresentam maior probabilidade de churn do que os que adquiriram banda larga mais recente. Quando a primeira banda larga tem tempo de base até 1 ano (até 365 dias), a chance de cancelamento é 1,3 vezes menor do que para aquela de mais de 3 anos (mais de 1.095 dias) de base.

Não há diferenças significativas entre não ter banda larga e ter banda larga com mais de 3 anos de base.

TEMPO DA PRIMEIRA B. LARGA	B	S.E.	Wald	df	Sig.	Exp(B)
> 1.095 dias			13	3	0,01	
Não tem banda larga	0,03	0,74	0	1	0,97	1,029
1 - 365 dias	(0,29)	0,08	12	1	0,00	0,751
366 - 1095 dias	(0,18)	0,08	5	1	0,03	0,837

Tabela 29 – Coeficientes da variável tempo da primeira banda larga.

Variável independente 13: **Tempo do primeiro móvel.**

Para a avaliação do tempo do primeiro móvel somente foi considerada a telefonia móvel pós-paga da empresa.

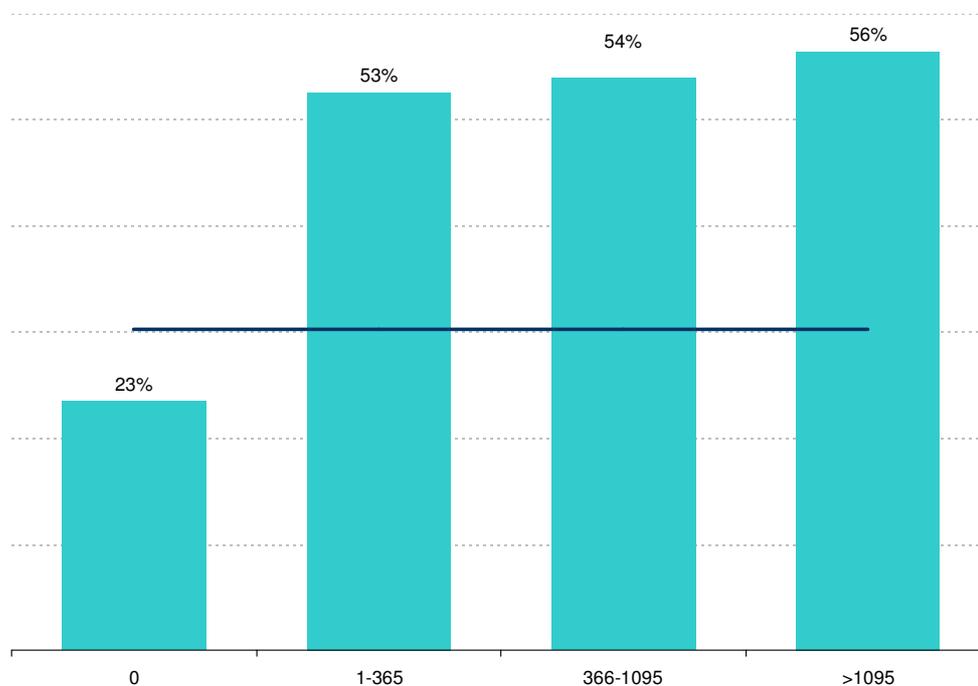


Gráfico 16 – Distribuição da amostra de *churn* por tempo do primeiro móvel.

Não há diferenças significativas entre não ter móvel ou ter o primeiro móvel de 1 a 3 anos (de 365 a 1.095 dias) de base, quando comparados aos que possuem móvel há mais de 3 anos.

Ter móvel até 1 ano de base apresenta 1,2 vezes menos chance de *churn* do que ter móvel há mais de 3 anos de base.

TEMPO DA PRIMEIRO MÓVEL	B	S.E.	Wald	df	Sig.	Exp(B)
> 1.095 dias			12	3	0,01	
Não tem móvel	0,68	0,68	1	1	0,32	1,975
1 - 365 dias	(0,18)	0,06	11	1	0,00	0,832
366 - 1095 dias	(0,06)	0,05	1	1	0,30	0,946

Tabela 30 – Coeficientes da variável tempo do primeiro móvel.

A tabela 31 apresenta uma consolidação dos resultados. Nela é possível observar que dentre as variáveis utilizadas, as comportamentais foram mais relevantes para a identificação do perfil do cliente propenso ao cancelamento. Das treze variáveis selecionadas pelo modelo, oito são comportamentais.

Variável	Descrição
1	Comportamental Quantidade de terminais fixos na residência A primeira variável selecionada pelo modelo foi a quantidade de terminais fixos na residência do cliente. Quanto maior a quantidade de terminais fixos maior a probabilidade de cancelamento.
2	Relacional Segmento Clientes novos e sem segmento são os que apresentam maior probabilidade de cancelamento. Entretanto, vale destacar que os clientes diamantes apresentam probabilidade de cancelamento maior que clientes bronze.
3	Comportamental Combinação de produtos Clientes que possuem outros produtos (móvel / móvel + banda larga), apresentam maior probabilidade de cancelamento do fixo, exceto para os casos de clientes que possuem o fixo + banda larga, que não apresentam diferenças significativas.
4	Comportamental Tempo de base do último fixo Quanto menor o tempo de base maior a probabilidade de cancelamento.
5	Comportamental Intenção anterior de cancelamento Cliente que já manifestou intenção anterior de cancelamento tem maior probabilidade de <i>churner</i> .
6.	Geográfica UF Espírito Santos, Rio de Janeiro, Sergipe e Bahia apresentam menor probabilidade de cancelamento.
7.	Característica do produto Disponibilidade de banda larga Regiões com menor disponibilidade de banda larga apresentam menor probabilidade de <i>churner</i> .

Tabela 31 – Tabela resumo das variáveis mais relevantes para o modelo.

	Variável	Descrição
8.	Demográfica	Idade do cliente
	Quanto menor a idade do cliente maior a probabilidade de cancelamento, desta forma, o público de 18 a 40 anos apresenta maior probabilidade ao <i>churner</i> .	
9.	Comportamental	Uso de ligação para celular
	Quanto menor a frequência de meses de utilização do fixo para efetuar ligação para celular, maior a probabilidade de cancelamento.	
10.	Mercado	Área de concorrência 1
	A atuação do principal concorrente influencia na probabilidade do churn. Onde há a atuação mais forte deste concorrente há uma probabilidade de cancelamento 1,3 vezes maior do que onde sua interferência é menor.	
11.	Comportamental	Uso de internet discada
	Clientes que utilizam internet discada apresentam probabilidade de cancelamento maior do que os que não utilizam.	
12.	Comportamental	Tempo de base da banda larga
	Clientes com banda larga com mais de 3 anos apresentam maior probabilidade de cancelamento do que aqueles que adquiriram banda larga mais recente.	
13.	Comportamental	Tempo de base do móvel pós-pago
	Clientes que possuem também o móvel pós-pago da mesma empresa há mais de 3 anos de base apresentam chances maiores de cancelamento.	

Tabela 31 (continuação) – Tabela resumo das variáveis mais relevantes para o modelo.

De forma simplificada, de acordo com as variáveis utilizadas no Modelo de Regressão Logística, o perfil dos clientes mais propensos a serem futuros *churners* é:

- possuir mais de um terminal;
- segmentos novos clientes ou sem segmentação;
- ter fixo + móvel ou fixo + móvel + banda larga;
- ter pouco tempo de base (principalmente se for segunda linha);
- ter tido intenção anterior de cancelamento e tiver sido retido;
- ter o terminal no RN, CE, PE ou PB;
- estar em região com disponibilidade de banda larga da empresa;
- clientes com idade de 18 a 40 anos;
- utilizar com menor frequência ligação de fixo para celular;

- estar localizado em região de forte atuação do concorrente 1;
- usar internet discada;
- possuir banda larga da empresa há mais de 3 anos; e
- possuir móvel pós-pago da empresa há mais de 3 anos.