

Churn Rate

Churn Rate

Depois de apresentar os dados de crescimento da empresa, queremos levantar o número de perdas de clientes. Como faremos para entender o comportamento de crescimento da sua empresa, do ponto de vista dos clientes. Sejam eles ativos, os que acessam pela primeira vez o site e o quanto temos de **churn** - taxa de clientes que abandonam a empresa.

Vamos tentar gerar alguns números que nos permitam pensar em estratégias. Precisamos ter estes tipos de números para poder contabilizar os clientes antes e depois de uma campanha de Marketing, por exemplo. Como ainda não temos os dados, teremos que calcular o *churn*.

Temos os dados do período de janeiro de 2014 até setembro de 2015, com os números de novos clientes e quantos estão ativos. Precisamos diferenciar o cliente que se cadastrou no site, daquele que está efetivamente trabalhando. Mas estas definições podem variar de uma empresa para outra. Um cliente pode ser considerado ativo se comprou nos últimos 30 dias. No caso da Jumping Cats, iremos considerar **cliente novo** aquele que se cadastrou no site ou na loja e como **ativo**, a pessoa que comprou um produto da loja.

A partir deste números, começaremos a construir quantas pessoas estão cancelando seu cadastro. Vamos ver o crescimento dos clientes ativos do mês de Fevereiro para Janeiro. A diferença será de 2.015 pessoas. No entanto, o número de cadastros foi superior: 2.602. A diferença entre os dois valores é 587. Ou seja, 587 clientes "sumiram". Na verdade, eles deixaram de comprar na Jumping Cats.

Para calcular o valor do *churn*, faremos uma conta simples: somaremos os valores dos clientes ativos de um determinado mês e depois, subtrairemos o número de novos clientes.

Meses	Novos Clientes	Clientes Ativos	587
1/1/2014	2,602	14,765	=C2+B2-C3
1/2/2014	3,177	16,780	
1/3/2014	3,374	19,122	
1/4/2014	2,742	20,633	

Vamos preencher todos os valores do *churn*.

Meses	Novos Clientes	Clientes Ativos	Churn(Clientes)
1/1/2014	2,602	14,765	587
1/2/2014	3,177	16,780	835
1/3/2014	3,374	19,122	1,863
1/4/2014	2,742	20,633	583
1/5/2014	3,418	22,792	1,797
1/6/2014	2,778	24,413	1,744
1/7/2014	2,804	25,447	1,440
1/8/2014	3,190	26,811	2,360
1/9/2014	3,000	27,641	2,918
1/10/2014	2,634	27,723	2,175
1/11/2014	2,514	28,182	1,970
1/12/2014	3,200	28,726	2,701
1/1/2015	2,884	29,225	1,350
1/2/2015	3,363	30,759	1,906
1/3/2015	3,220	32,216	3,326
1/4/2015	2,908	32,110	2,497
1/5/2015	2,616	32,521	3,072
1/6/2015	2,762	32,065	1,396
1/7/2015	3,276	33,431	3,046

Geralmente, o *churn* é calculado em percentual, ou seja, a quantidade de clientes relativo a base do valor anterior. Por isso, dividiremos a quantidade de clientes que saíram pelo número de clientes ativos do mês.

Clientes Ativos	Churn(Clientes)	Churn(%)
14,765	587	3.98%
16,780	835	4.98%
19,122	1,863	9.74%
20,633	583	2.83%
22,792	1,797	7.88%
24,413	1,744	7.14%
25,447	1,440	5.66%
26,811	2,360	8.80%
27,641	2,918	10.56%
27,723	2,175	7.85%
28,182	1,970	6.99%
28,726	2,701	9.40%
29,225	1,350	4.62%
30,759	1,906	6.20%
32,216	3,326	10.32%
32,110	2,497	7.78%
32,521	3,072	9.45%
32,065	1,396	4.35%
33,431	3,046	9.11%

Este é um cálculo de *churn* muito simples, existe várias outras formas de fazê-lo. Mas com os números que já levantamos, conseguimos ter uma noção do comportamento de cancelamentos dos clientes. Tivemos meses em que o *churn* foi maior e menor.

Empresas com mais tempo no mercado costumam ter um *churn* menor, enquanto empresas mais novas, costumam ter valores de *churn* maior. Então, é importante conhecer a empresa para analisar os dados.

Análise de Cohort

Outro tipo que veremos é a **análise de Cohort**, com a qual analisamos o comportamento de um grupo durante um período de tempo.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)
1/1/2014	1729
1/2/2014	1522
1/3/2014	1366
1/4/2014	1209
1/5/2014	972
1/6/2014	834
1/7/2014	750
1/8/2014	661
1/9/2014	530
1/10/2014	446
1/11/2014	362
1/12/2014	304
1/1/2015	263
1/2/2015	230
1/3/2015	193
1/4/2015	173
1/5/2015	147
1/6/2015	122
1/7/2015	107

Nós temos dados de clientes ativos de Janeiro de 2014 até Setembro de 2015. Imagine que você fez uma campanha de Marketing forte no rádio, em Janeiro de 2014. Vemos que entraram neste período, 1729 clientes. Destes, apenas 75 ficaram até Dezembro do ano seguinte.

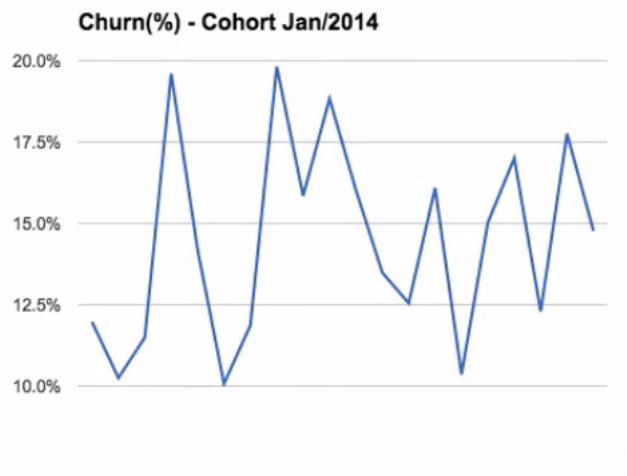
Para analisar esta queda de clientes, também faremos um levantamento percentual. Iremos subtrair os número de clientes de Janeiro pelo de Fevereiro, relativo ao número do primeiro mês.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)
1/1/2014	1729
1/2/2014	1522
1/3/2014	1366
1/4/2014	1209

Faremos uma análise *churn* dos demais meses.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)	Churn (%)
1/1/2014	1729	
1/2/2014	1522	12.0%
1/3/2014	1366	10.2%
1/4/2014	1209	11.5%
1/5/2014	972	19.6%
1/6/2014	834	14.2%
1/7/2014	750	10.1%
1/8/2014	661	11.9%
1/9/2014	530	19.8%
1/10/2014	446	15.8%
1/11/2014	362	18.8%
1/12/2014	304	16.0%
1/1/2015	263	13.5%
1/2/2015	230	12.5%
1/3/2015	193	16.1%
1/4/2015	173	10.4%
1/5/2015	147	15.0%
1/6/2015	122	17.0%
1/7/2015	107	12.3%

Podemos também criar um gráfico que daremos o título "Churn(%) - Cohort Jan/2014".



Se compararmos os valores do *churn* de agora, com os do exemplo anterior, vemos que os números de agora são superiores, variando de 10% a 20%. Na prática, quando temos uma campanha de Marketing muito forte é normal entrar muitas pessoas. Da mesma forma, é comum saírem muitas pessoas. Por isso, devemos ser cautelosos ao calcular o ROI (em inglês, *return on investment*) de uma campanha.

Continuando a análise de Cohort

Queremos agora calcular o *churn* médio e saber o tempo de permanência do cliente.

Como calcularemos a média? Precisaremos novamente calcular a **média geométrica**. Os dados que usaremos serão os seguintes:

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)
1/1/2014	1729
1/2/2014	1522
1/3/2014	1366
1/4/2014	1209
1/5/2014	972
1/6/2014	834
1/7/2014	750
1/8/2014	661
1/9/2014	530
1/10/2014	446
1/11/2014	362
1/12/2014	304
1/1/2015	263
1/2/2015	230
1/3/2015	193
1/4/2015	173
1/5/2015	147
1/6/2015	122
1/7/2015	107

Os 12% no inicio da lista é a perda de um mês para outro. Isto significa que os dados dos clientes ativos estão multiplicando 1 subtraído por 12%.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)
1/1/2014	1729
1/2/2014	1522
1/3/2014	1366
1/4/2014	1209

88.0% ×
=(1-C3)

De 1729 para 1522, tivemos uma perda de 12% de clientes. O número restante é 88% do anterior.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)
1/1/2014	1729
1/2/2014	1522
1/3/2014	1366
1/4/2014	1209
1/5/2014	972
1/6/2014	834
1/7/2014	750
1/8/2014	661
1/9/2014	530
1/10/2014	446
1/11/2014	362
1/12/2014	304
1/1/2015	263
1/2/2015	230
1/3/2015	193
1/4/2015	173
1/5/2015	147
1/6/2015	122
1/7/2015	107

Em seguida, iremos multiplicar todos os valores com a função `PRODUCT`. Iremos obter o número `0,04337767496`. Vamos calcular a raiz enésima, com truque já conhecido de usar a função `POWER`.

1/7/2015	107	12.3%	87.7%	
1/8/2015	88	17.8%	82.2%	
1/9/2015	75	14.8%	85.2%	
			0.8547976381	
			=POWER(D24, 1/20)	

Observe que usamos como potência $1/20$. A média dos valores será de $0,8547976381$, que em percentual é $85,48\%$. Se substituirmos o percentual anterior, pelo valor médio fixo, ao multiplicarmos pelo número de clientes ativos, teremos que chegar ao mesmo valor no fim.

1/3/2015	193	16.1%	83.9%	85.48%	192.2566552
1/4/2015	173	10.4%	89.6%	85.48%	164.3405348
1/5/2015	147	15.0%	85.0%	85.48%	140.477901
1/6/2015	122	17.0%	83.0%	85.48%	120.080178
1/7/2015	107	12.3%	87.7%	85.48%	102.6442525
1/8/2015	88	17.8%	82.2%	85.48%	87.74006461
1/9/2015	75	14.8%	85.2%	85.48%	75
		0.04337767496	85.48%		

Logo, o *churn* médio será de $85,48\%$.

Tempo médio

Encontramos o valor do *churn*. Agora, queremos identificar o **tempo médio** que um cliente permanece no grupo. Se analisarmos caso por caso a permanência de todas as pessoas e tirarmos uma média, qual valor encontrariamos? É uma pergunta difícil...

O grupo começou com uma quantidade grande e foi perdendo quase 15% de clientes por mês. Mas a medida que o número do grupo vai diminuindo, a tendência é que elas permaneçam mais.

Meses	Clientes Ativos (Cohort relativo à novos clientes de jan/2014)				
1/1/2014	1729				
1/2/2014	1522	12.0%	88.0%	85.48%	1477.945116
1/3/2014	1366	10.2%	89.8%	85.48%	1263.343995
1/4/2014	1209	11.5%	88.5%	85.48%	1079.903463
1/5/2014	972	19.6%	80.4%	85.48%	923.0989294
1/6/2014	834	14.2%	85.8%	85.48%	789.0627846
1/7/2014	750	10.1%	89.9%	85.48%	674.4890047
1/8/2014	661	11.9%	88.1%	85.48%	576.5516081
1/9/2014	530	19.8%	80.2%	85.48%	492.8349529
1/10/2014	446	15.8%	84.2%	85.48%	421.2741537
1/11/2014	362	18.8%	81.2%	85.48%	360.1041516
1/12/2014	304	16.0%	84.0%	85.48%	307.8161783
1/1/2015	263	13.5%	86.5%	85.48%	263.1205422
1/2/2015	230	12.5%	87.5%	85.48%	224.914818
1/3/2015	193	16.1%	83.9%	85.48%	192.2566552
1/4/2015	173	10.4%	89.6%	85.48%	164.3405348
1/5/2015	147	15.0%	85.0%	85.48%	140.477901
1/6/2015	122	17.0%	83.0%	85.48%	120.080178
1/7/2015	107	12.3%	87.7%	85.48%	102.6442525

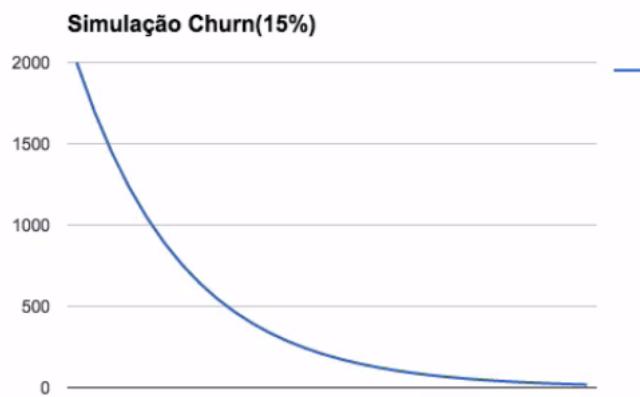
Vamos fazer um simulação, arredondando o valor inicial para 2000 pessoas.

Simulação	1700	×	2000
	=12*0.85		

Percebemos que a cada mês diminuirá a quantidade de pessoas que deixam o grupo.

Simulação	2000
	1700
	1445
	1228.25
	1044.0125
	887.410625
	754.2990313
	641.1541766
	544.9810501
	463.2338926
	393.7488087
	334.6864874
	284.4835143
	241.8109871
	205.5393391
	174.7084382
	148.5021725
	126.2268466
	107.2928196

Para visualizarmos melhor, vamos inserir um gráfico.



Analizando o gráfico, vemos que a quantidade de pessoas caiu, mas depois o número estabilizou e ficou próximo de 0. No entanto, a curva ainda deve seguir por muito tempo. Isto acontece porque os 15% estão sendo retirados de uma base cada vez menor.

Outra maneira de pensar o evento é pensar na situação em que todos os meses, algumas pessoas serão sorteadas aleatoriamente. Acompanhamos o período que ela permanecer até fazer *churn*. Faremos o mesmo com outras pessoas. O que acontecerá é que sempre teremos n eventos de *não churn*, seguido por um de *churn*. O estudo desta queda recebe o nome de **distribuição geométrica**. Por que é importante saber isto? A conclusão será: para respondermos a pergunta inicial "quanto tempo o cliente irá permanecer antes do *churn*?".

Em estatística, chamamos este caso de **esperança**, ou seja, quanto tempo **esperamos** que a pessoa fique antes do evento de *churn*. Podemos aplicar o conceito em outras situações, por exemplo, quantas vezes é preciso jogar antes de ganhar na loteria.

Se acessarmos o conceito de distribuição geométrica da [Wikipedia](#)

(https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_geom%C3%A9trica), veremos que o valor esperado é de $1/p$ sobre a probabilidade do evento (p):

Distribuição geométrica

Origem: Wikipédia, a enciclopédia livre.

Em [teoria das probabilidades](#) e [estatística](#), a **distribuição geométrica** é constituída por duas funções de probabilidade discretas:

- a distribuição de probabilidade do número X de [tentativas de Bernoulli](#) necessárias para alcançar um sucesso, suportadas pelo conjunto $\{1, 2, 3, \dots\}$, ou
- a distribuição de probabilidade do número $Y = X - 1$ de insucessos antes do primeiro sucesso, suportadas pelo conjunto $\{0, 1, 2, 3, \dots\}$.

Se a probabilidade de sucesso de cada tentativa é p , então a probabilidade de n tentativas serem necessárias para ocorrer um sucesso é

$$P(X = n) = (1 - p)^{n-1} p$$

para $n = 1, 2, 3, \dots$. De forma equivalente, a probabilidade de serem necessários n insucessos antes do primeiro sucesso é

$$P(Y = n) = (1 - p)^n p$$

para $n = 0, 1, 2, 3, \dots$.

Em qualquer caso, a sequência de probabilidades é uma [progressão geométrica](#).

Por exemplo, suponha um dado que é atirado repetidamente até à primeira vez que aparece um "1". A probabilidade de distribuição do número de vezes que o dado é atirado é suportado pelo conjunto infinito $\{1, 2, 3, \dots\}$ e é uma distribuição geométrica com $p = 1/6$.

O valor esperado de uma variável aleatória geometricamente distribuída X é $1/p$ e a variância é $(1 - p)/p^2$.

$$E(X) = \frac{1}{p}, \quad \text{var}(X) = \frac{1 - p}{p^2}.$$

De forma equivalente, o valor esperado de uma variável aleatória geometricamente distribuída Y é $(1 - p)/p$, e a sua variância é $(1 - p)/p^2$.

Como a sua distribuição contínua análoga (a [distribuição exponencial](#)), a distribuição geométrica tem a propriedade de perda de memória. Isto significa que se se tentar repetir uma experiência antes do primeiro sucesso, então, dado que o primeiro sucesso ainda não ocorreu, a função de distribuição condicional do número de tentativas adicionais não depende de quantos insucessos foram observados até então. A distribuição geométrica é, de facto, a única distribuição discreta com esta propriedade.

Isto significa que o **tempo médio** que podemos esperar que um cliente permaneça será 1 sobre a probabilidade, que é igual 14,52% .



O resultado será 6,886, o equivalente a 7 meses aproximadamente. Este tipo de informação é bastante útil, quando queremos calcular o ROI. Se um cliente gasta R\$ 1 mil por mês, sabemos que em média, ele irá render para a empresa R\$ 7 mil.

Este tipo de cálculo é bastante útil na parte financeira de uma empresa. É uma informação bastante relevante, que o seu chefe irá adorar receber.

