



Tempo médio

Transcrição

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 encontraríamos? 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

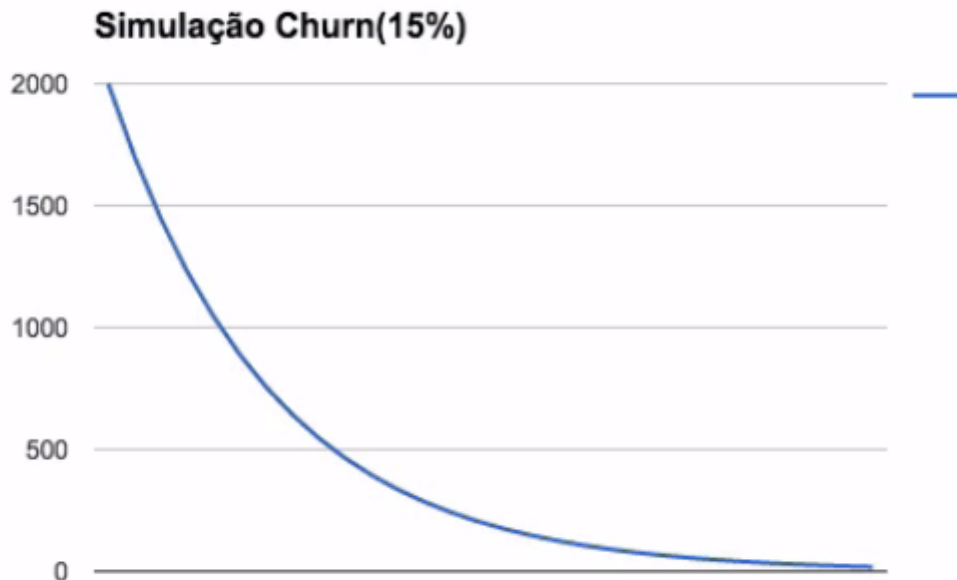
Vamos fazer um simulação, arredondando o valor inicial para 2000 pessoas.

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
	887.412625
	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) (https://pt.wikipedia.org/wiki/Distribui%C3%A7%C3%A3o_geom%C3%A9trica

veremos que o valor esperado é de 1 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 continua 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% .

	Churn Médio		6.886940316 x
0.04337767496	85.48%	14.52%	=1/F24

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.