

Análise Cohort

Transcrição

[00:00] Legal, pessoal, nós aprendemos um pouco sobre as métricas, sobre testes A/B e agora é hora de entender um pouco sobre cohort, análise cohort. O que é análise cohort? É uma análise de grupos, e basicamente, uma separação de subgrupos para nós podermos fazer uma análise mais assertiva.

[00:26] Quando nós trabalhamos com produtos digitais, esses produtos atingem uma quantidade de pessoas muito grande, e se nós tivermos sucesso, num espaço muito curto de tempo nós já conseguimos ter muitos usuários. E fazer essa análise, de muitos usuários, passa a ser algo complicado. Mesmo quando nós fazemos o teste A/B, como que nós vamos segmentar esse teste A/B? Será que vamos fazer direcionado? E se dentro de um grupo, vamos imaginar, nós fizemos lá aquela experiência de checkout, tirou o campo do cupom...

[01:02] Será que o comportamento de todos os nossos usuários vai ser igual? Será que o usuário que está comprando um produto mais caro vai ter um comportamento diferente do que o que está comprando mais barato? Ou que está comprando muito produto em relação ao que está comprando menos produto? Então, esse tipo de análise torna até esses testes um pouco mais ricos, porque quando nós temos muita gente é difícil de compreender esse pensamento coletivo, como que todas as pessoas funcionam.

[01:34] Então, nós podemos tentar dividir em grupos que têm algum tipo de semelhança num determinado espaço de tempo. Se fosse, por exemplo, vamos imaginar que nós temos um jogo online, temos um jogo que o pessoal está jogando e nós queremos fazer uma alteração nesse jogo para ver se as pessoas comprem mais itens lá no jogo. Nós ganhamos dinheiro vendendo item.

[02:00] Nós fazemos essa alteração, mas nós temos que começar a analisar, dos usuários que só criaram uma conta, os que criaram e jogaram partidas, pelo menos uma partida online e os que já compraram um item na loja. Nós temos alguns perfis um pouco diferentes. Vou falar da última semana, quem criou conta na última semana? Quem na última semana jogou uma partida online? E quem na última semana comprou algum item na loja? Já gastou dinheiro no produto?

[02:32] E nós fazemos a alteração e avaliamos cada um desses grupos de maneira isolada, porque muitas vezes uma ação, um experimento que nós fazemos, ele pode ter um impacto em um grupo e um impacto diferente em outro grupo. E separando dessa forma nós conseguimos compreender o comportamento dos nossos usuários um pouco melhor.

[02:53] Então, cada um desses grupos, na verdade, ele representa um grupo de usuários que eles participaram de um evento no mesmo período, como eu falei, quem criou a conta na semana passada, quem comprou um item na semana passada, nós agrupamos eles por algum evento do qual eles participaram numa determinada janela de tempo. Para nós fazermos esse tipo de análise, tem quatro passos que nós podemos tomar e fica fácil de entender mais ou menos como funciona.

[03:30] Primeira coisa, nós que determinar a pergunta, qual a pergunta que nós queremos responder? Nós vamos fazer um experimento, mas pra quê? O que nós queremos validar? E nós determinamos as métricas de avaliação. As métricas de avaliação, de preferência, nós vamos tentar bater ela com aqueles três "A". Nós vamos definir os grupos, quais são os grupos que vale a pena separarmos para estudarmos de maneira isolada, e iniciar a análise.

[04:01] Para ilustrar, eu vou citar um exemplo do livro do Dan Ariely, que é bem interessante. Ele tentou responder uma pergunta muito interessante, que é a origem da corrupção. E para fazer isso ele primeiro determinou a pergunta, será que eu consigo manipular as pessoas para elas terem comportamentos mais corruptos? Essa é a pergunta que ele queria responder. Então, ele pensou na métrica de avaliação, qual foi a métrica de avaliação dele? Ele criou um teste matemático, que era bem

difícil, e ele aplicou esse teste para um tanto de alunos de universidade e anotou a média de acertos que os alunos tinham. Estabeleceu uma média.

[04:52] A métrica de avaliação dele seria: para novos testes com situações diferentes que ele ia testar, o quão longe os novos grupos iam ficar dessa média. Então, se a média, vamos supor, eram sete acertos e, de repente ele muda uma condição e a coisa passa para 13, talvez as pessoas estivessem mentindo na resposta. O que ele fez foi assim, ele testou primeiro esse grupo, estabeleceu que essa seria a métrica de avaliação dele, e ele fez três grupos distintos.

[05:27] O primeiro grupo é o grupo que fez a prova naturalmente e entregou a prova. O segundo grupo, a pessoa ao invés de entregar ela anotava no papel quantos acertos ela, ela mesmo tinha um gabarito e falava: "eu acertei sete", ou "eu acertei oito", pegava a prova, passavam num triturador de papel para eliminar a prova. Então, ninguém tem como conferir se ela realmente acertou aquele número, se ela quiser mentir está tudo bem e ela entregava o resultado e ganhava um dólar por acerto.

[06:03] Esse era o segundo grupo. O terceiro grupo a pessoa também era bem semelhante a esse, mas a diferença é que na hora que ela contava quantos acertos ela obteve, ao invés de receber um dólar por acerto, a pessoa recebia uma fichinha de pôquer por acertos. E essa pessoa caminhava até o final do corredor e lá ela trocava cada ficha de pôquer por um dólar. Na prática é a mesma coisa, mas ela tem só uma etapa mais.

[06:33] E a pergunta que o Dan Ariely acabou tentando responder foi: será que quando eu afasto as pessoas do dinheiro elas se corrompem mais facilmente? E a resposta foi sim, então, ele tinha uma média de seis acertos, as pessoas que só contavam quantos acertos tiveram, mas pegavam dinheiro diretamente elas não mentiram tanto, a média saiu de seis foi para sete, é plausível. Mas as pessoas que pegavam a fichinha e depois trocavam fichinha pelo dinheiro, tiveram uma média de acerto de 9 a 10.

[07:10] Então, o que ele percebeu? Que quando as pessoas estão longe do dinheiro elas tendem a ter um comportamento um pouco mais corrupto. Por isso que às vezes, vamos supor, é mais fácil você deixar uma caneta em cima da mesa num ambiente público e alguém pegar a caneta, dentro de um escritório, onde as pessoas conhecem e essa caneta desapareceu, do que deixar uma nota de 10 reais, por exemplo, e essa nota sumir, porque nós temos a sensação que é pegar dinheiro. Foi essa pergunta que ele respondeu com essa pesquisa.

[07:48] E ele separou exatamente em pequenos grupos. O mundo de produtos digitais, se nós formos levar isso para cá, imagina: ao fazer essa análise nós conseguimos avaliar cada grupo. Nesse gráfico aqui do lado nós conseguimos ver quatro grupos, quatro linhas que representam ao longo do tempo a quantidade de usuários que pertence a cada um desses grupos, e nós conseguimos pontuar eventos que aconteceram que podem ter causado alguma alteração.

[08:21] Olha, quando nós mudamos o título de e-mail teve um pequeno ganho na confirmação por e-mail. Teve um degrau ali. Agora, quando nós implementamos o login mais simples, o login que precisa preencher usuário e senha, nós aumentamos bastante o número de usuário registrado. O número de login realizado também aumentou. O pessoal ficou com menos preguiça de usar o sistema.

[08:51] Esse tipo de análise é muito interessante para nós podermos fazer isso. Agora, é interessante perceber que correlação é diferente de causa, nós podemos ver que tem alguma relação ali, mas para afirmarmos que um evento causou alguma coisa, ou que quando uma curva sobe a outra sobe junto e desce a outra junto, nós temos que fazer uma análise um pouco mais aprofundada.

[09:16] Então, esse tipo de análise é muito interessante, por exemplo, voltando no exemplo do jogo é que eu comecei falando ali no começo. Imagina que nós temos lá esse jogo nosso, nós vendemos itens no mercadinho do jogo lá e nós separamos as pessoas que jogam em grupos, baseados na qualidade da conexão dessa pessoa. Uma pessoa que tem uma conexão muito boa, a pessoa tem uma conexão não tão boa e uma pessoa que tem uma conexão piorzinha. Para quem joga e conhece o termo, chama de lag. O atraso no tempo de resposta entre cliente e servidor.

[10:00] E nós começamos avaliar, no dia 1 nós tínhamos 10 usuários, então, dos 10 usuários que tinham um tempo de resposta bom, 9 fizeram compra. Dos 10 que tinham um tempo de resposta médio, oito fizeram compra. E dos 10 que tinham tempo de resposta ruim, 6 fizeram compras. Isso faz todo sentido, porque se ele está tendo uma experiência ruim no jogo, de repente está perdendo porque a conexão dele está ruim, ele não vai querer investir o tempo, o dinheiro dele naquilo.

[10:38] Esse é um exemplo de análise que se poderia fazer. Agora, é interessante sempre limitar pelo tempo. Nós limitamos pelo tempo. Naquele dia o que aconteceu, naquela semana o que aconteceu, nós separamos e estudamos esses grupos de maneira paralela, naquela linha de tempo, naquela janela de tempo. Vocês devem ter percebido que o segredo para fazer uma boa análise cohort é nós termos uma percepção boa dos subgrupos que existem dentro dos nossos clientes.

[11:16] Quando nós temos algum workflow, o cliente está na etapa X, na etapa Y, fica mais fácil de perceber esses grupos, eles já existem de uma maneira bem explícita, mas existem outros grupos, outros tipos de grupos que são um pouco mais difíceis de perceber, como esse da qualidade de conexão que mencionei. E cabe as pessoas que trabalham naquele produto tentar identificar esses grupos.

[11:45] Então, novamente, eu sei que eu já falei isso muitas vezes, é um trabalho para se fazer em grupo, porque quando nós temos pessoas diversas, cada um vai enxergar aquele conjunto de clientes sob um ângulo diferente. Essa pessoa vai enxergar agrupamentos distintos que vão poder enriquecer bastante esse tipo de análise.