

Descrição do Projeto

Você trabalha em uma startup que vende produtos alimentícios. Você precisa analisar o comportamento do usuário para o aplicativo da empresa.

Primeiro, estude o funil de vendas. Descubra como os usuários chegam à etapa de compra. usuários realmente chegam a essa etapa? Quantos ficam presos nas fases anteriores? Quais etapas em particular?

Em seguida, veja os resultados do teste A/A/B. (Continue lendo para obter mais informação sobre os testes A/A/B). Os designers gostariam de alterar as fontes de todo o aplicativo, mas os gerentes temem que os usuários achem o novo design intimidador. Eles decidem tomar a decisão com base nos resultados de um teste A/A/B.

Os usuários são divididos em três grupos: dois grupos de controle recebem as fontes antigas e um grupo de teste recebe as novas. Descubra qual conjunto de fontes produz melhores resultados.

A criação de dois grupos A tem certas vantagens. Podemos adaptar um princípio segundo o qual só estaremos confiantes na precisão de nossos testes quando os dois grupos de controle forem semelhantes. Se houver diferenças significativas entre os grupos A, isso pode nos ajudar a descobrir fatores que podem distorcer os resultados. A comparação de grupos de controle também nos informa de quanto tempo e dados precisaremos ao executar outros testes.

Você usará o mesmo conjunto de dados para análise geral e análise A/A/B. Em projetos reais, os experimentos se realizam constantemente. Os analistas estudam a qualidade de um aplicativo usando dados gerais, sem prestar atenção na participação dos usuários em experimentos.

Descrição dos dados

Cada linha do registro (também referido como log) é uma ação do usuário ou um evento.

- `EventName` — nome do evento
- `DeviceIDHash` — identificador de usuário exclusivo

- `EventTimestamp` — hora do evento
- `ExpId` — número do experimento: 246 e 247 são os grupos de controle, 248 é o grupo de teste

Instruções para completar o projeto

Passo 1. Abra o arquivo de dados e leia a informação geral

Caminho do arquivo: `/datasets/logs_exp_us.csv` [Baixe o conjunto de dados](#)

Passo 2. Prepare os dados para análise

- Renomeie as colunas de uma maneira que seja conveniente para você
- Verifique se há valores ausentes e tipos de dados. Corrija os dados, se necessário
- Adicione uma coluna de data e hora e uma coluna separada para datas

Passo 3. Estude e verifique os dados

- Quantos eventos ficam nos registros?
- Quantos usuários ficam nos registros?
- Qual é o número médio de eventos por usuário?
- Qual é o período de tempo que os dados cobrem? Encontre as datas máxima e mínima. Desenhe um histograma por data e hora. Você pode ter certeza de que possui os dados igualmente completos para todo o período? Os eventos mais antigos podem acabar aparecendo nos diários de alguns usuários por motivos técnicos, e isso pode distorcer o quadro geral. Encontre o momento em que os dados começam a ser completos e ignore a seção anterior. Qual período os dados realmente representam?
- Você perdeu muitos eventos e usuários ao excluir os dados mais antigos?
- Certifique-se de ter usuários de todos os três grupos experimentais.

Passo 4. Estude o funil de eventos

- Veja quais eventos estão nos diários e sua frequência de ocorrência. Classifique-os por frequência.
- Encontre o número de usuários que executaram cada uma dessas ações. Ordene os eventos pelo número de usuários. Calcule a proporção de usuários que executaram a ação pelo menos uma vez.
- Em que ordem você acha que as ações ocorreram? Todas elas fazem parte de uma única sequência? Você não precisa levá-las em consideração ao calcular o

funil.

- Use o funil de eventos para encontrar a parcela de usuários que passam de uma etapa para a próxima (por exemplo, para a sequência de eventos $A \rightarrow B \rightarrow C$, calcule a proporção de usuários na etapa B para o número de usuários na etapa A e a proporção de usuários na etapa C para o número na etapa B).
- Em qual fase você perde mais usuários?
- Qual é a parcela de usuários que faz o caminho inteiro, desde o primeiro evento até o pagamento?

Passo 5. Estude os resultados do experimento

- Quantos usuários há em cada grupo?
- Temos dois grupos de controle no teste A/A, no qual verificamos nossos mecanismos e cálculos. Veja se há uma diferença estatisticamente significativa entre as amostragens 246 e 247.
- Selecione o evento mais popular. Em cada um dos grupos de controle, encontre o número de usuários que realizaram essa ação. Encontre a proporção deles. Verifique se a diferença entre os grupos é estatisticamente significativa. Repita o procedimento para todos os outros eventos (economizará tempo se você criar uma função especial para este teste). Você pode confirmar se os grupos foram divididos corretamente?
- Faça a mesma coisa para o grupo com fontes alteradas. Compare os resultados com os de cada um dos grupos de controle para cada evento isoladamente. Compare os resultados com os resultados combinados para os grupos de controle. Quais conclusões você pode tirar do experimento?
- Qual nível de significância você definiu para testar as hipóteses estatísticas mencionadas acima? Calcule quantos testes de hipóteses estatísticas você realizou. Com um nível de significância estatística de 0,1, um de cada 10 resultados pode ser falso. Qual deveria ser o nível de significância? Se você quiser alterá-lo, execute as etapas anteriores novamente e verifique suas conclusões.

Ao concluir este projeto, você irá provar suas habilidades:



Ou faça o projeto em seu computador e faça upload de seu trabalho quando finalizar.

Fazer upload e e... 

Como você achou do projeto?  