

Missão Stone

Estudo de rentabilidade

A planilha em anexo **Base_Missão_S.xlsx** contém dados fictícios de alguns clientes ao longo de 3 meses. Na primeira aba, você vai encontrar um dataset com o volume transacionado, a receita transacional e a receita de antecipação, assim como o ID dos clientes e seus segmentos de atuação. Na segunda aba, você vai encontrar o ID dos clientes e seu custo de aquisição, o CAC, contabilizado apenas 1 vez na vida útil do cliente.

O intuito do estudo é que você explore os dados e responda as seguintes perguntas:

1. Qual seria sua sugestão de clusterização da base? Por quê?
2. Estando em um cenário de concorrência acirrada:
 - a. Qual estratégia comercial poderíamos implementar para diminuir nossa perda de clientes?
 - b. Quais clientes você priorizaria para blindá-los de um eventual ataque da concorrência e por quê? Não é necessário especificar os IDs dos clientes, apenas explique quais são as características dos clientes e quantos são.
3. Qual sua sugestão para maximizar o lucro da companhia?

Dica: Caso sinta falta de alguma informação necessária para a conclusão do case utilizando sua linha de raciocínio, estime-a e explicito o racional.

Monte uma **apresentação** mostrando sua linha de raciocínio e recomendações. Pedimos que também nos envie suas análises em excel, python, etc em anexo até as 23:59 do dia anterior a apresentação.

Estudo de dados

Você tem acesso a dois datasets, “*Transactions.csv*” e “*Clients.csv*”. Esses arquivos contêm, respectivamente, informações transacionais (para o período entre jan/21 e dez/22) e cadastrais de clientes. Abaixo, segue uma descrição das variáveis presentes nesses datasets:

Transactions:

client : Código identificador do cliente.
date : Mês de referência.
sales_volume : Volume de vendas (em R\$).
sales_count : Quantidade de transações.

Clients:

client : Código identificador do cliente.
document_type : Tipo de documento cadastrado.
created_at : Data de credenciamento.
segment : Segmento de atividade econômica.

O intuito do estudo é que você explore os dados e responda as seguintes perguntas:

1. Para começar, você decidiu calcular o volume médio de vendas de cada cliente, para classificá-los dentro de diferentes intervalos de faturamento.
 - a. Apresente 3 sugestões de metodologia de classificação. Na sua opinião, quais são as vantagens e desvantagens de cada uma das definições? Qual é a que você usaria na prática? Por quê?
 - b. Utilizando a sua definição predileta, crie uma nova variável, tier, que determine o intervalo de faturamento médio de cada cliente. Explique a escolha da quantidade de intervalos adotados. Dica: Utilize abreviações como K = R\$1 mil e M = R\$1 milhão, por exemplo tier "100K-1M".
2. Considerando as variáveis que você já tem à sua disposição (bem como outras que você julgue relevante criar), produza estatísticas descritivas, incluindo tabelas e plots, que na sua avaliação podem trazer insights interessantes para gerar valor ao negócio.
3. Uma métrica importante para acompanhamento na indústria de meios de pagamento é o churn da base de clientes. Tipicamente, a taxa de churn é calculada pela razão entre a quantidade de clientes sem transacionar em um determinado mês (mas ativos no mês anterior) e a quantidade total de clientes que estavam ativos (i.e. transacionando) no mês anterior. Considerando a definição de churn acima, crie uma variável indicadora churn para identificar se o cliente está ativo (churn = 0) ou deu churn (churn = 1) em cada date e crie um plot com a evolução da taxa de churn da companhia e por tier ao longo do tempo. Quais insights você destacaria para as lideranças da companhia?

Para esse estudo, fique à vontade para mostrar seu raciocínio diretamente em ferramentas como o Jupyter Notebook. Pedimos que também nos envie suas análises em anexo até as 23:59 do dia anterior a apresentação.