

Framework de avaliação - Heitor Felix de Oliveira

a. Quantos contratos distintos existem em sua tabela de observação final, após todos os merges e etapas de ETL?

14.756 contratos distintos

b. Você considerou ou aplicou algum filtro de exclusão? Caso afirmativo, descreva-o e informe quantos contratos foram filtrados.

Alguns filtros foram aplicados no portfólio geral e no portfólio tpv. Todos valores negativos foram apagados. Mas nenhum contrato foi excluído, somente alguns registros com valores estranhos.

- Linhas removidas para valores de pagamento negativo: 23
- Linhas removidas para valores de pagamento negativo: 1591
- Linhas removidas para valores de quantidade de transação negativo: 7
- Linhas removidas para valores de transação de transação negativo: 1238

Na tabela com contratos únicos foram desconsiderados linhas de contrato duplicadas com valor de tipo de empresas diferentes, devido ao pouco impacto que teria.

Etapas 1 - Limpeza dos dados

- Para a tabela *portfolio_geral* foram selecionadas somente as colunas que variam e chaves primárias (contrato_id (PK), dt_ref_portfolio (PK), nr_documento, status_contrato, perc_retencao, vlr_pgto_realizado, vlr_pgto_esperado, vlr_saldo_devedor, vlr_saldo_devedor_esperado, dsp, dspp, flag_transacao, qtd_transacoes, vlr_tpv). A tabela filtrada passou a ser chamada de *df_geral*;
- As colunas que eram constantes, foram separadas em outra tabela chamada *df_contrato*. Os duplicados foram removidos e esse dataframe contém todas as informações do contrato, chaves primárias e estrangeiras (contrato_id (PK), nr_documento (FK), safra, dt_contrato, dt_desembolso, dt_vencimento, dt_wo, prazo, vlr_desembolsado, vlr_tarifa, juros_mes, juros_diario);
- Foram removidas todas as linhas que tinham *status_contrato* = 'Settled' exceto a primeira, pois uma vez que o empréstimo foi quitado, os próximos registros não interessam. Cerca de 25% das linhas foram removidas aqui, sem prejuízo de informação;

- Valores negativos de quantidade e valor foram removidos (considerando o total, pouca informação foi removida). Tanto nas tabelas *df_geral* e *portfolio_tpv*.

Objetivo: A limpeza foi feita principalmente para reduzir o número de linhas e melhorar a performance do script e do dashboard e também para excluir valores esquisitos que poderiam afetar a análise.

Etapa 2 - Join de tabelas

- Join da tabela *df_geral* com o *portfolio_comunicados* pelo *contrato_id* e *dt_ref_portfolio*, para extrair os comunicados (se houve) em cada registro histórico do contrato (tipo left join). Onde não houve comunicação, ficou com valores de ação nulos e alguns registros foram duplicados pois houveram mais de uma ação no mesmo dia para o mesmo contrato, foi criado portanto uma nova tabela chamada *df_geral_comunicado* que recebeu esse join;
- Join da tabela *df_contrato* com o *portfolio_clientes* pelo *nr_documento*, para extrair as informações do cliente de cada contrato (tipo inner join). Essa foi armazenada no próprio *df_contrato*;
- Join da tabela *df_geral* com *portfolio_tpv* pelo *nr_documento* e *dt_transacao* com *dt_ref_portfolio*, para extrair as informações de transação pela maquininha em cada registro histórico do contrato (tipo left join), valores nulos foram substituídos por 0, pois significa que houve 0 transação.

Objetivo: Conseguir cruzar informações e correlacionar variáveis que estavam separadas.

Etapa 3 - Criação de novas tabelas

- As colunas que eram constantes, foram separadas em outro dataframe chamado *df_contrato*. Os duplicados foram removidos e esse dataframe contém todas as informações do contrato. Algumas linhas tinham valores de *contrato_id* duplicado com informações de tipo de empresa diferente, mas como eram poucos casos, os duplicados foram removidos;
- Uma tabela chamada *df_mensagem* foi criada apenas com as linhas do *df_geral_comunicado* que continha uma comunicação e os 5 registros seguintes. O objetivo desse dataframe é observar o efeito da comunicação nos 5 dias seguintes, através do *dsp* e *dspp*. Dessa vez o enfoque é descobrir a curva ideal de vezes que se deve acionar o cliente.

Objetivo: A tabela de contratos teve o objetivo de sumarizar todas as informações referentes a cada contrato, além de resumir as informações que se repetiam no portfólio geral. Já a tabela de mensagem teve o objetivo de analisar os acionamentos e a eficiência dos comunicados.

Etapa 4 - Criação de novas features

4.1. Geral

- Valor pago t_{pv} = Valor (t_{pv}) * percentual de retenção
- Valor debitado sem pagamento - Valor que do débito aumentando conforme os juros, sem considerar o pagamento

4.2 Contrato

- Última data - Última data registrada pelo contrato;
- Status final do contrato - Qual o status do contrato no último dia registrado;
- Média de dsp do contrato - média aritmética de dsp, extraída do *df_geral*;

- Média de dspp do contrato - média aritmética de dspp, extraída do *df_geral*;
- Soma do valor pago tpv - soma do valor pago tpv, extraída do *df_geral*;
- Soma do valor pago realizado - soma do valor pago realizado, extraído do *df_geral*;
- Valor total pago - Soma do valor pago tpv + Soma do valor pago realizado;
- Valor debitado desconsiderando pagamento - Máximo do valor debitado, que corresponde ao total que o cliente pagaria no final do contrato;
- Tempo esperado - Tempo em dias entre data do desembolso e data do vencimento do contrato;
- Dias de contrato - Tempo em dias entre primeira data e a última de contrato;
- Nº de ações totais - Número total de ações enviada registradas;
- Nº de ações entregues - Número total de ações , desconsiderando as que não foram entregues;
- Média de eficiência - média aritmética da eficiência, extraída do *df_mensagem*;
- Lucro esperado - $d\acute{v}ida_original * ((1 + j\acute{u}ros_diario)^{dias_de_contrato}) - 1$).

4.3. Mensagem

- Eficiência - (1) se o comunicado foi lido e o cliente pagou algo nos próximos 5 dias (dsp), (2) se o comunicado foi lido e o cliente pagou o principal nos próximos 5 dias (dspp), (0) comunicado foi lido e o cliente não pagou nos próximos 5 dias. Os valores partem da premissa de que a eficiência é maior se há amortização do saldo principal.

Objetivo: Extrair informação do conjunto de dados original para fazer as análises. As features referente às ações e sua eficiência foram chave para conseguir encontrar a resposta para a pergunta do case.

Etapas 5 - Criação do Dashboard Power BI

5.1. Análise temporal

- **Filtro de ano:** Filtro obrigatório pois as análises são divididas por ano.
- **Cartões:** Informações gerais de número de contratos assinados, valor total desembolsado, valor total pago e número de contratos ativos (aqueles onde o último registro é do ano analisado).
- **Gráfico de área - Pagamento total mensal, realizado e esperado:**
Compara o valor de pagamento que era esperado para aquele mês do ano com o valor efetivamente pago.
- **Tabela** - Essa tabela mostra valores acumulados de pagamento e desembolso, mostrando a evolução mensal tanto do custo do programa quanto do retorno.
- **Gráfico de linha** - Número de contratos quitados, realizado e esperado por trimestre. Os contratos quitados são aqueles que possuem status “quitado” e o último registro é no trimestre do ano analisado, os esperados são aqueles em que o vencimento é no trimestre do ano analisado.

Principais insights:

- Em 2020: O pagamento realizado superou o esperado por quase todo o ano. Mas os contratos foram sendo quitados numa frequência menor do que o esperado.
- Em 2021: O pagamento realizado recuou bastante em relação ao esperado, mas o pagamento acumulado desde o início começou a chegar perto do valor total desembolsado. O ritmo de quitação dos contratos também superou o esperado para esse ano.
- Em 2022: O pagamento mensal não está tão legal, mas o acumulado provavelmente vai alcançar o valor desembolsado esse ano.

5.2. Impacto das ações

- **Cartões:** Informações gerais de eficiência média, média de dsp, média de dspp, número total de comunicados enviados e número total de comunicados entregues (status “entregue”, “lido” e “respondido”).

- **Gráfico de linha - Média de eficiência por média de dsp:** este gráfico mostra a relação da eficiência ao longo do tempo sem pagar. A conclusão é de que quanto mais a pessoa fica sem pagar, menos eficientes são as ações em média. As linhas pontilhadas no eixo x mostram onde estão concentradas as ações de acordo com a régua de acionamento.
- **Gráfico de barras horizontais - Média de eficiência por ação:** Este gráfico compara a eficiência média de cada campanha.
- **Gráfico de funil - Funil de status das ações:** Esse gráfico mostra a proporção do total de mensagens que são entregues, lidas, respondidas e não entregues.
- **Filtros:** Este é um filtro que consegue mostrar a relação entre duração de um contrato, tipo de empresa, segmento, subsegmento, estado e cidade na eficiência média das ações.

Principais insights:

1. O comportamento médio geral da eficiência e os dias sem pagar são muito variáveis;
2. A eficiência média das ações tendem a cair bastante a medida que o contrato se prolonga;

5.3. Análise individual por contrato

- **Gráfico de linha - Eficiência por dsp:** este gráfico mostra a relação da eficiência ao longo do tempo sem pagar. Dessa vez é possível observar a correlação individual para cada contrato, quando o filtro dinâmico é ativado. Sem utilizar médias, é possível perceber o quanto esse comportamento é variável.
- **Tabela** - Essa tabela o número de ações lidas e respondidas de cada campanha e suas respectivas eficiências médias. As células de tipo de ação são pintadas de azul, com um espectro variando com base na eficiência média (quanto mais eficiente, mais azul).
- **Filtros:** Este é um filtro que consegue mostrar a relação entre tipo de empresa, segmento, subsegmento, estado e cidade na eficiência média das ações.

- **Play Axis - filtro dinâmico:** Este é um filtro adicional que tem a função de alterar de forma automática um filtro. Neste caso, o filtro seleciona cada *contrato_id*. Esses contratos estão ordenados pelo número de ações enviadas para conseguirmos ver primeiro o efeito naqueles que mais foram acionados. Este é o principal diferencial em relação a página anterior.

Principal insight: O comportamento em relação ao tipo de ação, e a eficiência das ações ao longo do tempo sem pagar é muito variável. Não é possível perceber uma grande linearidade de comportamento. Mas com certeza é possível aferir que o melhor intervalo de acionamento até o dia 10 sem pagamento, depois disso o comportamento varia muito.

O dashboard está disponível neste [link](#):

Etapas 6 - Análise de dados em Python

Todas as análises estão detalhadas no notebooks, mas estes são os principais insights:

- O valor que ainda está para ser pago é: R\$81,450,085.45;
- O saldo até agora foi negativo com o valor de : R\$-11,707,166.48;
- 27.85% dos contratos estão atrasados;
- Taxa de leitura das ações: 35.09%;
- Taxa de eficiência das ações lidas: 42.16% ;
- No script é possível observar o impacto que teria caso mais mensagens fossem entregues;
- A campanha de Pré Negativação é mais eficiente que a campanha de Boleto Quitado, mesmo sendo uma campanha mais tardia.

Conclusão: qual é a curva ideal de vezes que devemos acionar um cliente?

Sobre a eficiência do sistema, foi detectado um problema de alta taxa de comunicados não entregues. De todas as mensagens enviadas, cerca de 53% são entregues, portanto, muito do esforço do envio das mensagens é desperdiçado por falhas no sistema. Além disso, em relação às mensagens que são entregues, somente 35% das mensagens são lidas e respondidas.

Agora para entender a eficiência de cada ação, precisamos metrificar a eficiência de cada ação. O primeiro dado importante é que 42% das mensagens surtiram efeito e as pessoas voltaram a pagar. Mas destrinchando a eficiência média das ações, conclui-se que as campanhas mais eficientes são de observação (0,72), parcelamento (0,43) e pré-negativação (0,27).

Naturalmente as primeiras ações surtem maior efeito, mas a conclusão interessante foi que a pré negativação foi mais eficiente que a campanha de boleto quitado (0,21). Com base nesse estudo, deve-se garantir que estas campanhas devam continuar com a estratégia atual e a campanha de boleto quitado pode ser melhorada.

Por fim, é importante entender que as ações funcionam muito mais com pouco tempo de atraso no pagamento. Até os 10 primeiros dias sem pagamento, deve haver um foco maior no envio de comunicados. Garantindo que os comunicados cheguem nesse período. Outro fator interessante é perceber que a eficiência média das campanhas são maiores para contratos de menor duração.