

Ficha Técnica do Projeto

Projeto: Segmentação de Clientes via RFM – O Mercado (Projeto 1)

Analista de Dados: Bruna Derner

Ferramentas utilizadas: Google Sheets, Looker Studio, Google Apresentações e Loom.

1. Objetivo do Projeto

A loja “O Mercado” busca compreender melhor o comportamento da sua base de clientes a fim de aumentar as fidelizações – desafio enfrentado num cenário altamente competitivo e com mudanças nos padrões de preferência do consumidor. Sendo assim, este projeto tem como objetivo realizar uma segmentação de clientes usando a metodologia RFM (Recência, Frequência e Valor Monetário), fornecendo informações que permitam tomada de decisões estratégicas em marketing, fidelização e retenção.

2. Google Sheets

2.1. Processamento e Preparação dos Dados

2.1.1. Importação dos dados.

Os dados foram importados diretamente das planilhas de origem utilizando a função IMPORTRANGE combinada com QUERY para filtrar registros nulos e outliers.

Usando a fórmula (exemplo):

=QUERY(

IMPORTRANGE("https://docs.google.com/spreadsheets/d/1qK0a1k8dd_MhEff2mArOvi7de61FS-3hPnRyTC7Wus4/edit?usp=sharing"; "clientes!A1:I2241");

"select * where Col1 is not null and Col2 is not null and Col3 is not null and Col4 is not null and Col5 is not null and Col5 <> 666666 and Col6 is not null and Col7 is not null and Col8 is not null and Col9 is not null";

1
)

2.1.2 Tratamento de Dados Nulos e Outliers

Antes da importação dos dados foi analisado quais dados eram nulos, quais eram outliers e eu tomei a decisão de tirá-los da análise ao invés de tratá-los.

Esses foram os dados encontrados a partir do CONTAR.VAZIO e estatística da coluna, ferramenta do próprio google sheets:

Clientes: 24 valores nulos removidos (principalmente em salário)

Transações: 7 valores nulos removidos

Resumo de compras: Nenhum valor nulo encontrado

Outlier: Um valor de salário muito alto (666666) foi excluído no momento da importação

2.1.3 Tratamento de Duplicados

9 registros duplicados de id_cliente identificados na tabela resumo_compras e removidos (eram cópias idênticas), identificados a partir da ferramenta de limpeza de dados duplicados do google sheets.

2.1.4 Unificação das Tabelas

As tabelas foram unidas usando FILTER e CORRESP para garantir que apenas os IDs presentes nas 3 bases fossem mantidos:

```
=FILTER(  
  transacoes!A:D;  
  ÉNÚM(CORRESP(transacoes!B:B; clientes!A:A; 0));  
  ÉNÚM(CORRESP(transacoes!B:B; resumo_compras!A:A; 0))  
)
```

Em seguida, foi utilizado PROCV para cruzar as informações das tabelas buscando em cada uma das três bases de dados.

2.1.4 Criação de Variáveis Derivadas

Variáveis criadas:

idade: =2022 - E2

```
faixa_etária: =IFS(  
  2022 - E2 < 25; "18 a 24";  
  2022 - E2 < 35; "25 a 34";  
  2022 - E2 < 45; "35 a 44";  
  2022 - E2 < 55; "45 a 54";  
  2022 - E2 < 65; "55 a 64";  
  VERDADEIRO; "65 ou mais"
```

)

periodo_compras: =TEXTTO(C2; "yyyy-mm") para analisar a sazonalidade

faixa_salarial (: =SE(H2 < 30000;

"Até 30k";

SE(H2 < 60000;

"30k - 60k";

SE(H2 < 90000;

"60k - 90k";

"Acima de 90k"

)

)

)

filhos (com classificação): =IFS(I2

+J2 = 0; "Sem filhos";I2

+J2 = 1; "1 filho";I2

+J2 = 2; "2 filhos";

VERDADEIRO; "3 ou mais filhos"

)

total_gasto: =SOMA(M2:R2)

frequencia: =CONT.SES(A:B;B2)

ticket_médio (total gasto / número de compras): =SE(Y2 > 0; SOMA(M2:R2) / Y2; 0)

A de total_gasto e frequência eu só fiz para calcular ticket médio. Mas posteriormente fiz novamente da de RFM para que fosse contabilizada de forma assertiva.

2.2 Análise Exploratória Inicial

Ao terminar a criação de variáveis foi realizada uma análise inicial para ter alguns insights. Foram criadas tabelas dinâmicas e gráficos que apresentaram os seguintes resultados:

Nível de escolaridade dos clientes

A maioria dos clientes da loja possui ensino superior completo (50,3%), seguido de pós-graduação (38,3%). Apenas 11,4% têm apenas o nível secundário.

Isso indica que os clientes são escolarizados e possivelmente são mais exigentes e possuem renda mais elevada.

Estado civil

Casados são o maior grupo (38,7%)

Em seguida, temos união estável (25,8%) e solteiros (21,5%)

Divorciados, viúvos e "outros" representam menor parte

Esses resultados podem indicar que as decisões de compra são mais familiares.

Faixa salarial

R\$ 30k - 60k é a faixa mais comum (45,5%)

Depois, R\$ 60k - 90k (36,3%)

Faixas extremas representam menos da metade (acima de R\$ 90k: 2,2%, até R\$ 30k: 16%)

Os resultados indicam que a maior parte dos clientes tem um bom poder aquisitivo e boa capacidade para consumo.

Faixa etária

O maior grupo está na faixa dos 45 a 54 anos(32,7%)

Seguido por 55 a 64 anos (21,8%) e 65 anos ou mais (20,8%)

O público da empresa é um público mais maduro, provavelmente com mais estabilidade financeira e que valoriza a qualidade e confiança.

Análise de Consumo de Produto

O produto mais consumido disparado é o vinho, com 52,3% do total gasto

Em seguida, carnes (26,1%)

Outros produtos (peixes, doces, frutas e "outros") aparecem de forma secundária.

Se vinho é o carro chefe da empresa talvez o marketing devesse focar neste produto oferecendo promoções e artigos de fidelização.

Além dessas informações também foi possível verificar onde se concentra a maior parte das compras, que é em loja (58,68%) frente a compras on-line (41,32%).

Com isso podemos concluir essa primeira análise traçando um perfil de clientes.

Escolaridade em nível superior (nível superior/pós)

Maduro (acima de 45 anos)

Com renda de média a alta

Casado ou em união estável

Preferencialmente realiza compras na loja física

E tem uma forte preferência por vinhos frente a outros produtos.

2.3 Segmentação RFM

Para fazer a análise de RFM as informações não podem ficar repetidas, sendo assim foi necessário criar uma nova tabela para poder classificar os clientes usando o quartil.

Na nova tabela foram adicionadas todas as variáveis tidas como importante, começando pelo id_clientes =UNIQUE(Compilado!B:B). Depois foi usado PROCV para adicionar escolaridade, estado civil, faixa etária, número de filhos, faixa salarial e resposta da campanha.

Para calcular os quartis utilizei a fórmula =QUARTIL(I2:I2206;0) para todas as variáveis e todos os quartis.

Quartil	recencia1	frequencia1	valor_monetari o1
0	827	1	8
1	881	5	69
2	960	10	397
3	1070	14	1048
4	1706	27	2525

Posteriormente foi utilizado a fórmula =IFS(i2<=1; 1; i2<=5; 2; i2<=10; 3; i2<=14; 4; i2<=27; 5) para classificar cada cliente. Os valores variam de 1 a 5, cada cliente pode obter até 5 pontos em recência, frequência e valor monetário. Chegamos a classificação:

Clientes fiéis: são clientes recentes, frequentes e que gastam bem: $R \leq 2$; $F \geq 4$; $M \geq 4$

Fiéis em potencial: recentes, com frequência boa e valor razoável: $R \leq 3$; $F \geq 3$; $M \geq 3$

Promissores: recentes, mas ainda não compram muito: $R \leq 3$; $F \leq 3$; $M \leq 3$

Clientes em risco: inativos, mas costumavam comprar bem: $R \geq 4$; $F \geq 3$; $M \geq 3$

Clientes perdidos: inativos e com baixo histórico de compras: $R \geq 4$; $F \leq 2$; $M \leq 3$

Grandes gastadores: gastam muito ($M \geq 4$), mesmo sem serem tão recentes: $M \geq 4$; $F \geq 3$; $R \geq 3$

Clientes novos: muito recentes, ainda não compraram muito: $R \leq 2$; $F \leq 2$; $M \geq 2$

Clientes em desenvolvimento: quase lá — perfil mediano em tudo: R entre 2 e 5; F entre 2 e 4; M entre 2 e 4

Clientes estáveis: média geral, sem destaque relevante: não se encaixam nas categorias anteriores (perfil por volta de 3)

Utilizamos a fórmula:=IFS(

E(R2>=4; T2>=4; S2>=3); "Grandes gastadores";

E(R2<=2; S2<=2; T2>=2); "Clientes novos";

E(R2<=2; S2>=4; T2>=4); "Clientes fiéis";

E(R2<=3; S2>=3; T2>=3); "Fiéis em potencial";

E(R2<=3; S2<=3; T2<=3); "Promissores";

E(R2>=4; S2>=3; T2>=3); "Clientes em risco";

E(R2>=4; S2<=2; T2<=3); "Clientes perdidos";

E(R2>=2; R2<=5; S2>=2; S2<=4; T2>=2; T2<=4); "Clientes em desenvolvimento";

VERDADEIRO; "Clientes estáveis"

)

3. Looker Studio

A importação de dados foi feita via planilhas google. Foi realizada a união das tabelas RFM e Compilado para que o dashboard ficasse mais interativo, a mescla escolhida foi externo pela direita para os dados sobre transação ficarem armazenados em cada id de clientes.

Dessa forma, o dashboard ficou mais automatizado e podemos verificar todas as informações por segmento do cliente. Foram inseridos gráficos de sazonalidade e de perfil do cliente, scorecards com resultados absolutos e tabelas com a segmentação e compra por categoria de produto, além de um filtro de classificação de cliente.

4. Resultados e conclusões

Indicadores Gerais:

Foram inseridos 4 scorecards para os indicadores:

Total de transações: 21.902

Total de clientes: 2.205

Receita total: R\$1.339.929

Ticket-médio geral: R\$35

Sazonalidade e transações ao longo do tempo:

O gráfico de área mostra um crescimento consistente até meados de 2022, seguido por uma leve queda no final do período.

Perfil de Clientes:

Escolaridade:

Distribuição dos níveis educacionais

Graduação 50,2%

Pós-graduação 38,2%

Ensino Secundário 11,5%

Isso indica um público altamente escolarizado, que tende a:

Preferir informações detalhadas sobre os produtos;

Responder bem a campanhas com apelo racional

Estado civil

Casados são a maioria

União estável e solteiros: seguem na sequência

Maior foco em casais e famílias pode melhorar o engajamento.

Faixa Salarial:

A maioria dos clientes está concentrada em "30k - 60k" (45,5%) e "60k - 90k" (35,7%).

Clientes com esse perfil tendem a buscar custo-benefício: qualidade com preço justo;

Produtos muito sofisticados ou com preços excessivos podem ter baixa adesão se não forem bem posicionados.

Filhos:

58,9% dos clientes não têm filhos

A maioria dos compradores têm mais liberdade financeira.

Faixa Etária:

O maior grupo está na faixa dos 45 a 54 anos(32,7%)

Seguido por 55 a 64 anos (21,8%) e 65 anos ou mais (20,8%)

O público da empresa é um público mais maduro, provavelmente com mais estabilidade financeira e que valoriza a qualidade e confiança.

Bom, de acordo com o que foi encontrado é possível concluir que a análise baseada na metodologia RFM permitiu segmentar a base de clientes da loja O Mercado e identificar perfis com comportamentos e potenciais distintos. Observou-se que três segmentos concentram a maior parte da receita, destacando-se os **Fiéis em Potencial**, **Clientes Fiéis** e **Grandes Gastadores** — que juntos representam mais de 64% da base de clientes e mais de 95% de toda a receita da empresa.

Com base nesses achados, elaborou-se recomendações estratégicas organizadas em quatro frentes prioritárias:

Foco Estratégico: direcionado a **Clientes Fiéis**, **Grandes Gastadores** e **Fiéis em Potencial** que são os segmentos mais valiosos e com forte impacto na receita, merecendo programas de fidelização e ações de retenção imediata.

Potencial de Crescimento: abrange **Clientes Novos** e **Clientes em Desenvolvimento** que são clientes que ainda não atingiram todo seu potencial, mas que respondem bem a estímulos como campanhas sazonais e vantagens exclusivas.

Observação: voltado para **Clientes Promissores**, que ainda não atingiram seu pleno valor. Esses perfis exigem incentivos moderados e um plano de acompanhamento próximo.

Última Chance: contempla **Clientes em Risco** e **Clientes Perdidos**, com sinais claros de inatividade. Para eles, recomenda-se estratégias de reativação imediata ou, se necessário, a realocação de esforços para outros segmentos mais rentáveis.

5. Limitações\ Próximos passos:

Limitações: A análise foi realizada com base em dados históricos e comportamentais disponíveis até o momento, o que pode não refletir mudanças recentes nos hábitos dos consumidores.

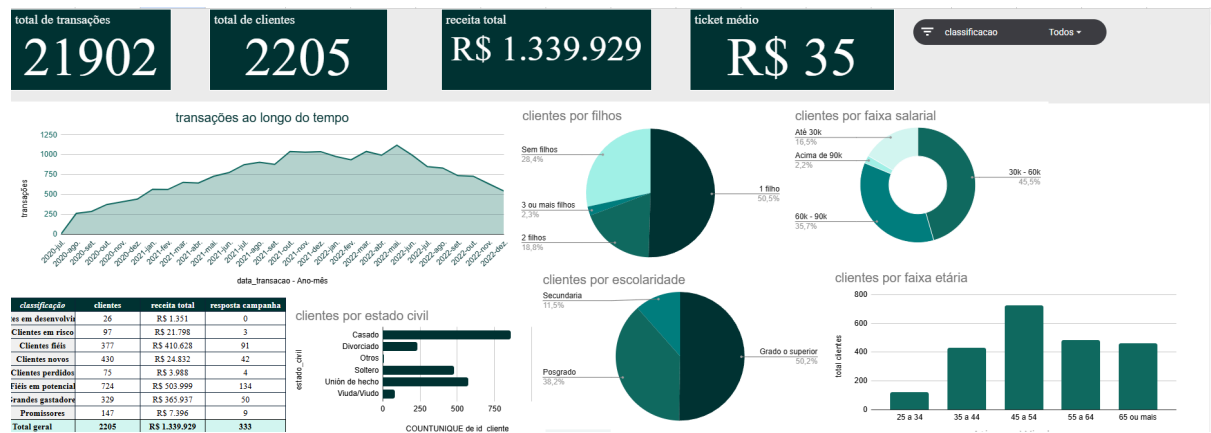
Próximos Passos: Eu apostaria na utilização de outras ferramentas para análise para que ela ficasse mais rica, não utilizaria apenas RFM. Utilizaria K-means etc, mas para isso teria de usar outras ferramentas também.

Faria também o monitoramento da resposta às ações implementadas e atualizaria a segmentação periodicamente, garantindo que as estratégias estejam alinhadas com a evolução da base de clientes.

6. Links de Interesse

☐ Google Sheets

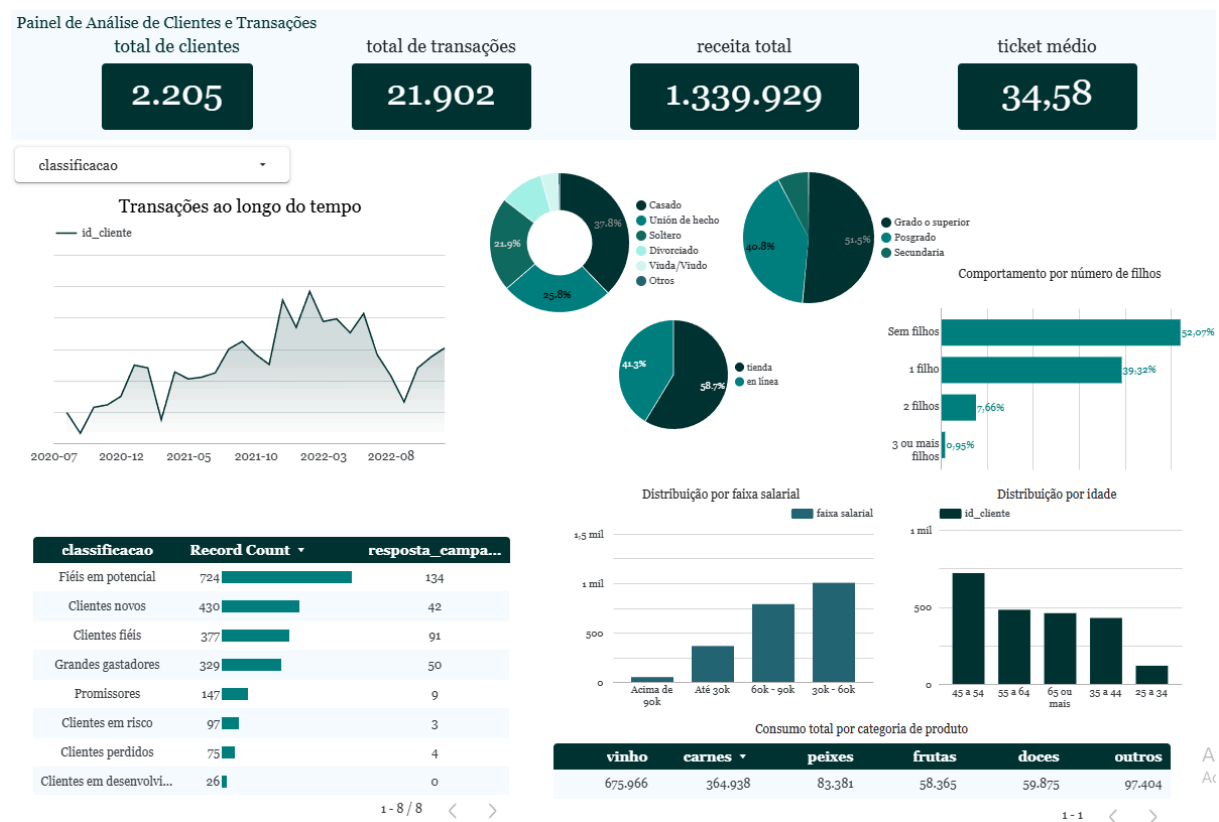
Link da planilha com o dashboard no Google Sheets:
https://docs.google.com/spreadsheets/d/1_E2wwfR9YxtgV2mbxb2xPde1LLIZ5F9Zm2hjZDqOISQ/edit?usp=sharing



Google Sheets

☐ Looker Studio

Link do dashboard no Looker Studio:
<https://lookerstudio.google.com/reporting/ff9d9626-041a-447c-a0ff-dbbe6034c224>



Looker Studio

☐ Google Apresentações

Link da apresentação no Google Apresentações:

https://docs.google.com/presentation/d/1Fprojiw9jptm_NRDVWqwtumcAwOZAW2724u4OoHQ3yg/edit?usp=sharing