# **Processamento e análises:**

# **Processamento e preparação da base de dados**

* Usando a fórmula: IMPORTRANGE importei os dados para uma única planilha.
* Identifiquei com a fórmula CONT.VALORES() 24 campos em branco na aba clientes, na coluna de salario\_anual\_dolar, optei por excluir esses dados .
* Depois utilizei a formatação condicional junto com a fórmula ==[CONT.SE](http://cont.se/)($A:$A, A1)>1 para procurar dados duplicados, identifiquei que na aba copia\_resumo\_compras haviam 09 linhas com dados duplicados, excluir esses dados para não ocorrer problemas na análise.
* Na aba de transações utilizei a formula **CONT.VALORES()** e encontrei inconsistências na coluna id\_cliente onde havia 07 linhas de clientes sem cadastro, efetuei a exclusão desses clientes que não havia cadastro.
* Usei a função PROCV para unir as informações das abas clientes e resumo\_compras, com o nome da aba (dados\_integrados) e Usei =[CONT.SE](http://cont.se/)(transacoes!B:B;A2) para criar uma nova coluna com a soma das transações por id\_cliente nessa mesma aba.
* Criei as variáveis abaixo na aba dados\_integrados\_rfm  
  1. Idade usando a fórmula =2022 -B2 na aba
  2. faixa\_etaria utlizando: =SE

1. total\_filhos utilizando: =SOMA
2. ano\_entrada utilizando: =ANO
3. mes\_entrada utilizando = TEXTO(K2;"mmmm")
4. dia\_entrada utlizando: =TEXTO(K2;"dddd")
5. total\_gastos utilizando: =SOMA
6. ano\_ultima\_transacao: =ANO
7. mes\_ultima\_transacao: =TEXTO(X2;"MMMM")
8. dia\_ultima\_transacao: = TEXTO(K2;"dddd")
9. dias\_sem\_compras: =HOJE()

* Usei: classificar página para deixar os id\_cliente em ordem, crescente.
* Usei na aba dados\_integrados na coluna idade a função FILTRO e encontrei 3 idades que considerei fora do escopo, pois eram idades de 123, 122 e 128 anos, considerei que foi um erro de digitação na hora de passar essas informações para a planilha, optei por excluir esses dados para não ocorrer interferências na análise.
* Na aba de transacoes\_limpo criei a variável abaixo:
  1. Ultima\_transacao
* Na aba dados\_integrados\_rfm agrupei as informações e criei tabelas dinâmicas em uma nova aba chamada: tabelas\_dinamicas e criei algumas variáveis categóricas e alguns gráficos para visualizar melhor as informações.
* Na aba transações agrupei os dados e criei tabelas dinâmicas e criei algumas variáveis categóricas e em seguida criei gráficos para uma melhor visualização dos dados.
* Na aba dados\_integrados calcule os quartil Q1, Q2, Q3 e Q4, de recência, frequência e quantia, utilizando a fórmula: =quartil.
* Nessa mesma aba criei novas colunas chamadas recência, frequência e quantia e calculei o rfm utilizando a fórmula: IF e criei uma nova coluna chamada rfm\_score para calcular a média do rfm utilizando: =SOMA.
* Na coluna chamada rfm\_segmentação criei a segmentação por cliente utilizando a fórmula: IF com as variáveis.

1.Campeão

2.Cliente fiel

3.Cliente necessita atenção

4.Cliente promissor

5.Clientes potenciais leais

6.Hibernando

7.Não perder

8.Novos clientes

9.Potencial perda

10.Quase dormindo

* Com os dados da rfm criei uma tabela dinâmica na mesma aba com as informações de rfm\_segmentacao e a conta dos clientes por categoria.
* Em uma nova aba chamada: Dashboard criei um resumo com gráficos e as informações mais relevantes

Ao total no Marco 1 foram excluidos os seguintes ID\_CLIENTE 8720,1994,5376,11181,11110,5255,7829,7281,7244,8557,10629,8996,11004,9235,5798,8268,1295,8475,2437,2863,5555,3955,10749,1150,6862,10475,2902,4345,3769,4931,9931,7187,1612,5079,10339,3117.

Marco 2:

utilizei =PROCV para puxar a coluna de data\_entrada da aba dados\_integrados\_rfm.

Na coluna F utilizei =CONCATENAR(ANO(F2);"/";TEXTO(MÊS(F2);"00")) para agrupar ano/mês da data\_entrada e na coluna H também utilizei porém para agrupar ano/mês da coluna C data\_transacao.

* Criei a com a fórmula =DATADIF(H2;G2;"M") a variável para saber quantos meses se passaram desde a data de entrada a data de transação.
* Calculei MOD, MED E MEDIA.
* E com a variável meses\_que\_passaram criei um histograma.
* Apliquei análise de Cohorte criando uma tabela dinâmica com as variáveis id\_cliente, ano\_mes\_entrada e meses\_que\_passaram.
* E criei um mapa de calor para obter uma melhor visualização da análise de cohorte.
* Criei a taxa de retenção ulizando a fórmula = =(B3/B$27), alterando para celula subsequente.

Marco 3:

Integrei a tabela com as colunas no Looker Studio e criei um dashboard com todas as informações que creio serem relevantes para a analise, link:<https://lookerstudio.google.com/u/1/reporting/37621377-12cb-4b53-8a81-f0c9cb5d1a2b/page/zyxuD/edit>

E utilizando o Canvas criei uma apresentação com todos os dados que achei relevante com minhas recomendações.