**Ficha técnica- Projeto 1 (Segmentação)**

**Link para a planilha do projeto:**

https://docs.google.com/spreadsheets/d/1XhVNaa84BnGD0Rnov9xCuvaNp-NGtZkPEDmdA8e6jko/edit?usp=sharing

**Início**

* Leitura de todo o projeto
* Criação de cópias das planilhas para manter as originais

**1- Processar e preparar a base de dados**

A- Conectar/importar dados

* Uso da fórmula IMPORTRANGE para importar as três planilhas.

B-Identificar e tratar valores nulos

* Os valores nulos foram identificados com a fórmula COUNTBLANK
* Foram identificados poucos valores nulos nas tabelas "clientes" e transacoes" comparados ao total de clientes, portanto foram excluídos da planilha base
* As planilhas "clientes" e "transacoes" foram importadas novamente já sem os valores nulos com a fórmula QUERY + IMPORTRANGE (abas "clientes\_sem\_nulos" e "transacoes\_sem\_nulos")
* Não foram identificados valores nulos na tabela "resumo\_compras"

C- Identificar e tratar valores duplicados

* Foram identificados poucos valores duplicados apenas na tabela "resumo\_compras", com formatação condicional (fórmula =CONT.SE(A:A;A1)>1)
* Os valores duplicados foram excluídos manualmente da planilha base e importados novamente com a fórmula IMPORTRANGE (aba "resumo\_compras\_sem\_duplicados)

D- Identificar e gerenciar dados fora do escopo da análise

* Foram identificados apenas na análise exploratória

E- Unir tabelas

* Como a tabela de transações continha várias linhas para cada id de cliente , foi criada uma nova aba (resumo\_transacoes) com uma tabela dinâmica resumindo quantas compras cada cliente fez online, na loja e o total de compras
* Para identificar apenas os ids presentes nas três planilhas foi criada a aba ids\_verificados com uma tabela contendo uma coluna para ids dos clientes, uma para registro de transações e uma para registro de compras. Para identificar os ids presentes na aba "resumo\_transacoes" foi usada a fórmula SE+CONT.SE, para que retornasse "positivo" caso o id estivesse presente e "negativo" caso não estivesse presente. O mesmo foi feito para a aba "resumo\_de\_compras\_sem\_duplicados"
* Para unir as tabelas na aba "tabelas\_unidas" foi criada uma coluna para os ids, importando os ids da aba "ids\_verificados" com a fórmula QUERY (=QUERY(ids\_verificados!A2:C2214;"select Col1 where Col2= 'positivo' and Col3= 'positivo'")
* As outras colunas das abas "clientes\_sem\_nulos" e "resumo\_compras\_sem\_dupicados" foram importadas com a fórmula PROCX
* As colunas da aba "resumo\_transacoes" foram importadas com a fórmula SE + PROCX para substituir valores vazios por zero

F- Criar novas variáveis

* Foi criada uma coluna para dividir os clientes por faixas etárias (18 a 24 anos, 25 a 34 anos, 35 a 44 anos, 45 a 54 anos, 55 a 64 anos e 65 anos ou mais) usando a fórmula SE

**2- Fazer uma análise exploratória**

A- Agrupar dados de acordo com variáveis categóricas/ B- Visualizar variáveis categóricas

* Foi criada a aba "variaveis\_categoricas" para que fossem criadas tabelas dinâmicas de cada variável categórica e a partir delas gráficos de barra para melhor visualização. Com esses gráficos ficou evidente que a maior parte dos clientes tem entre 45 a 54 anos, escolaridade de nível superior, são casados e que poucos responderam à campanha
* Com os gráficos foram identificados três clientes com mais de 100 anos e excluídos da planilha base (Identificar e gerenciar dados fora do escopo da análise)

C- Calcular quartil, decil ou percentil

* Leitura dos conteúdos indicados para aplicar na próxima fase do projeto

**3-Aplicar técnica de análise**

A- Aplicar segmentação

* Foi criada a aba "rfm" para aplicação da segmentação
* Tabela com as variáveis do RFM:
  + RECÊNCIA: Para descobrir os valores de recência foi criada a aba "data\_ultima\_compra" onde foi inserida uma tabela dinâmica (feita a partir da aba "resumo\_transacoes") para se ter a data da última compra por id. Essas datas foram incluídas em uma coluna na aba "tabelas\_unidas" e a partir dela foi criada outra coluna para contar os dias transcorridas desde a última compra subtraindo-se da fórmula HOJE a data da última compra. Essa coluna de dias desde e última compra colada na aba da RFM.
  + FREQUÊNCIA: A coluna "soma\_online\_loja" foi copiada da aba "tabela\_unidas"
  + VALOR: Foi criada uma coluna com a soma dos produtos na aba "tabelas\_unidas" que depois foi copiada e colada na aba da RFM
* Tabelas dos quintis: Foram calculados os valores dos quintis de cada variável RFM com a fórmula PERCENTIL e foram atribuídas notas de 1 a 5 para cada uma.
* As notas da RFM foram atribuídas a cada cliente a partir da tabela dos quintis, com a fórmula SE
* Foi usada como base uma tabela com os onze segmentos da RFM e a partir dela a segmentação foi aplicada aos clientes com a fórmula SE