

1. Descrição dos Dados

Foram utilizados quatro conjuntos de dados:

- **Customer.csv**: contém informações dos clientes, como nome, gênero, data de nascimento e data de criação no sistema.
- **Orders.csv**: contém informações sobre os pedidos, incluindo datas (criação, atualização, agendamento), canal de venda, tipo de pedido, status, valor total, tempo de preparação e tempo de retirada.
- **CampaignQueue.csv**: contém informações de campanhas enviadas aos clientes, incluindo identificadores (jobId, campaignId, storeId, customerId), telefone, datas de agendamento, envio e atualização, status, mensagens enviadas e respostas recebidas.
- **Campaign.csv**: contém informações das campanhas de marketing cadastradas, incluindo id, nome, tipo, status, descrição, datas de criação/atualização e atributos relacionados à execução.

Após a padronização dos nomes das colunas, foram criadas variáveis derivadas como:

- Idade e Faixa Etária dos clientes (16-20, 21-25, 26-30, 31-45 e 46+).
 - Mês de criação dos pedidos e clientes para análises temporais.
 - Junção (merge) entre clientes e pedidos pelo identificador do cliente, permitindo análises cruzadas.
 - Indicadores de engajamento de campanhas: taxa de resposta, quantidade de mensagens por cliente, top clientes mais acionados.
 - Indicadores de performance de campanhas: distribuição por tipo (Promocional x Institucional), status (Draft, Published, Completed, Canceled) e evolução temporal de criação/envio.
-

2. Exploração dos Dados

Customer e Orders

- **Distribuição temporal:** clientes e pedidos foram agrupados por mês, identificando períodos de maior atividade.
- **Perfil dos clientes:** análise por gênero e faixa etária, incluindo distribuição de idades.
- **Pedidos:** distribuição por status, tipo de pedido e canal de venda.
- **Indicadores de negócio:** ticket médio por tipo de pedido, status, canal de venda e faixa etária.
- **Eficiência operacional:** cálculo do tempo médio de preparação e retirada, analisado por canal de venda e faixa etária.
- **Análises interativas:** dashboards com Plotly/Dash permitem filtros dinâmicos (por faixa etária, gênero e canal).

CampaignQueue

- **Distribuição dos status:** variando de 1 a 6, indicando diferentes estados das campanhas.
- **Respostas dos clientes:** apenas 32% das mensagens tiveram resposta, mostrando baixo engajamento.
- **Envios ao longo do tempo:** houve períodos de maior concentração de mensagens, analisados por data de envio.
- **Clientes mais acionados:** alguns clientes receberam significativamente mais mensagens (top 10 clientes com maiores volumes).
- **Taxa de resposta por cliente:** variou bastante, com alguns respondendo mais de 70% das vezes, enquanto a maioria respondeu pouco ou nada.

Campaign

- **Distribuição por Status:** campanhas agrupadas em Draft, Published, Completed e Canceled. Observou-se que a maioria chega a Published/Completed, mas há volume considerável de cancelamentos.
 - **Distribuição por Tipo:** predomínio de campanhas **Promocionais**, indicando maior foco em vendas do que em comunicação institucional.
 - **Evolução Temporal:** picos de criação/envio em meses com datas comemorativas, sugerindo planejamento sazonal. O tempo entre `createdAt` e `sendAt` varia, mostrando diferenças de execução.
 - **Campanhas em Destaque:** algumas campanhas aparecem com maior recorrência, sugerindo priorização de esforços de marketing.
-

3. Verificação da Qualidade dos Dados

Customer e Orders

- **Valores nulos:** em campos de clientes (ex.: gênero, data de nascimento) e pedidos (ex.: datas e atributos operacionais).
- **Datas inválidas:** algumas entradas de `dateOfBirth` precisaram ser tratadas com `errors="coerce"`.
- **Padronização de colunas:** necessidade de remover espaços e ajustar padrões para facilitar a análise.
- **Distribuição desigual:** algumas faixas etárias e canais de venda têm menos registros, prejudicando representatividade.

CampaignQueue

- **Valores nulos:** `sendAt` ausente em 34% dos casos e `response` ausente em 68%.

- **Telefone (phoneNumber):** problema de formatação (números com casas decimais), necessitando correção.
- **Datas:** campos `scheduledAt`, `sendAt`, `createdAt` e `updatedAt` estavam como texto e precisaram de conversão para datetime.
- **Distribuição de status:** códigos de 1 a 6 sem documentação explícita, exigindo validação com a fonte de dados.

Campaign

- **Valores nulos:** campos de descrição e datas ausentes em algumas campanhas.
 - **Datas incoerentes:** registros em que `scheduledAt` ou `sendAt` ocorrem antes de `createdAt`.
 - **Consistência:** enums de status e tipo seguem o padrão esperado, mas há concentração em poucos valores.
-

4. Conclusão

A análise exploratória dos quatro conjuntos de dados permitiu compreender melhor o perfil dos clientes, comportamento dos pedidos e desempenho das campanhas.

- **Customer/Orders:** fornecem insights de perfil por faixa etária, eficiência operacional e impacto do canal de venda.
- **CampaignQueue:** revelou baixa taxa de resposta e concentração de campanhas em poucos clientes, levantando pontos sobre saturação e segmentação.
- **Campaign:** mostrou predominância de campanhas promocionais, evidência de sazonalidade e taxas relevantes de cancelamento.

Esses resultados podem orientar decisões estratégicas em três frentes:

1. **Marketing:** melhorar segmentação e reduzir saturação de clientes com alto volume de disparos.
2. **Operações:** alinhar tempos de preparo/retirada com canais de venda.
3. **Gestão de campanhas:** equilibrar esforços entre campanhas promocionais e institucionais, além de revisar causas de cancelamentos.