

Processamento Analítico de Dados

Data Warehousing para Análise de CRM

- Gabriel Ribeiro Fonseca de Freitas - 12542651
- Henrique Souza Marques - 11815722
- Laura Fernandes Camargos - 13692334
- Letícia Barbanera Menezes - 14588642

Introdução: CRM para E-Commerce



<https://www.centripe.ai/>

O que é CRM?

- Estratégia + tecnologia para **gerenciar interações com clientes**.

Relevância:

- Visão 360° do cliente (histórico, preferências, contatos).
- Melhora da **experiência do cliente** e da **fidelização**.
- Centraliza dados de vendas, marketing e suporte.
- Base para decisões orientadas a dados.

Estrutura do DW

Medallion
Architecture

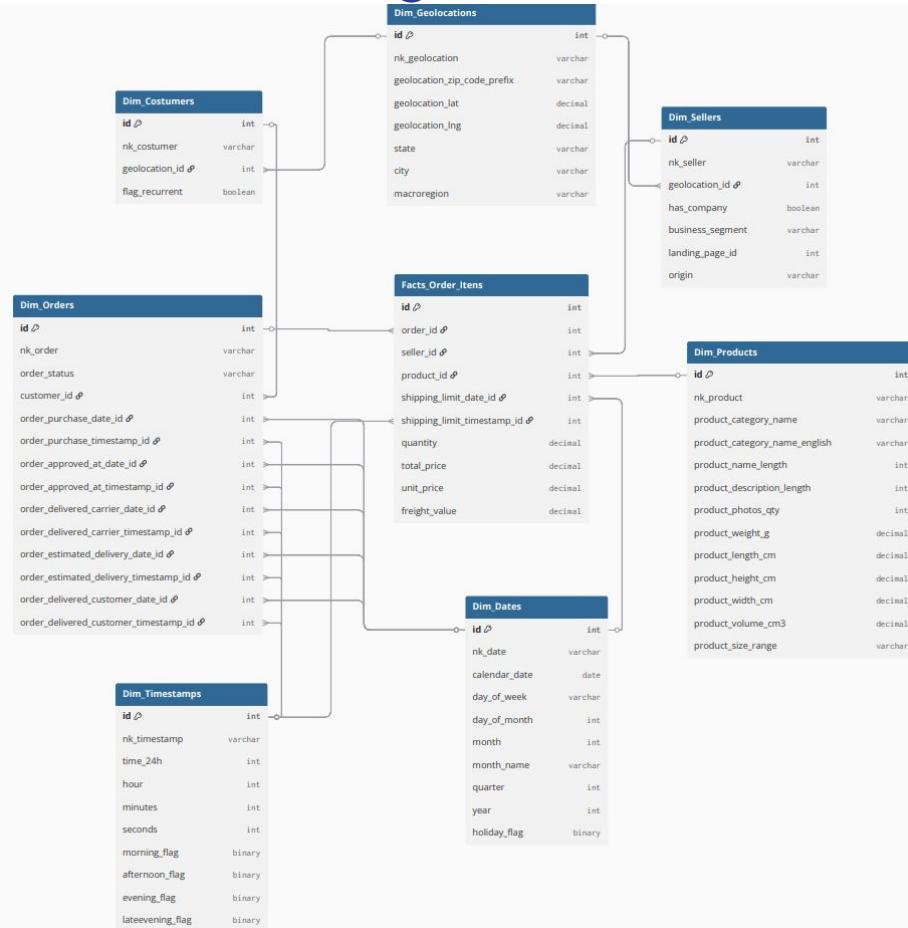


APACHE
Spark™

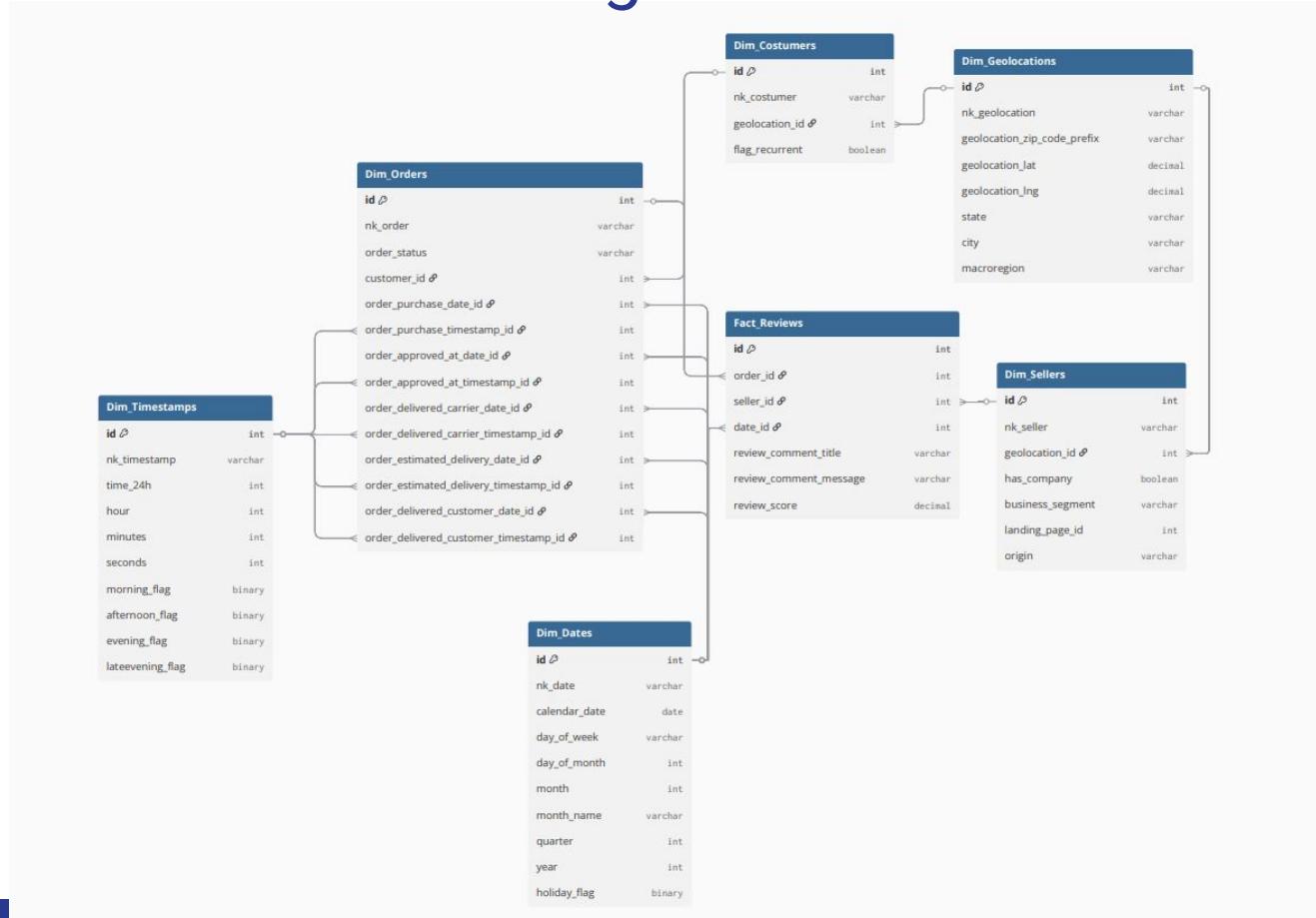
Mas antes:
Relembrando...
Nosso Modelo Dimensional



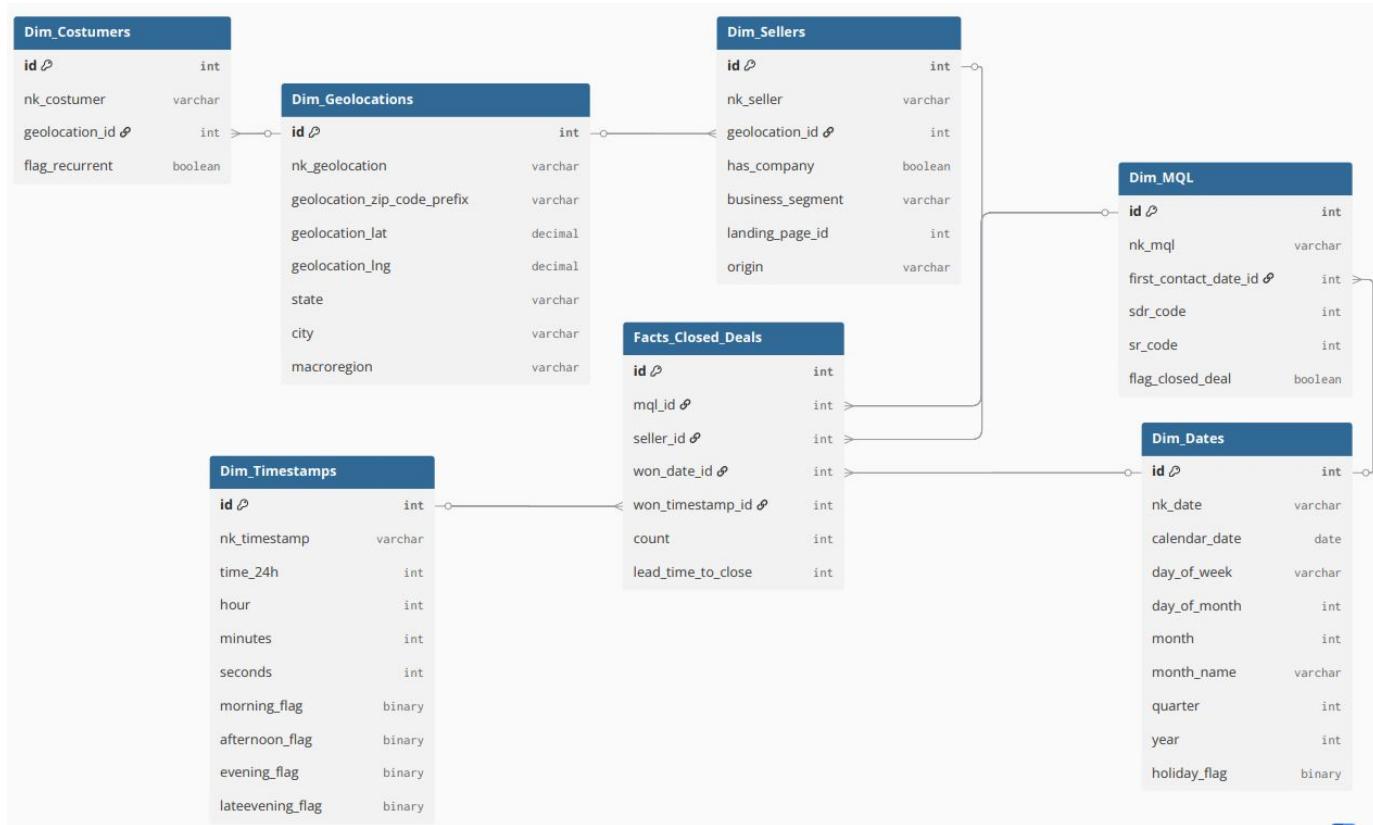
Modelo Dimensional Corrigido



Modelo Dimensional Corrigido



Modelo Dimensional Corrigido



Consultas - Drill Down

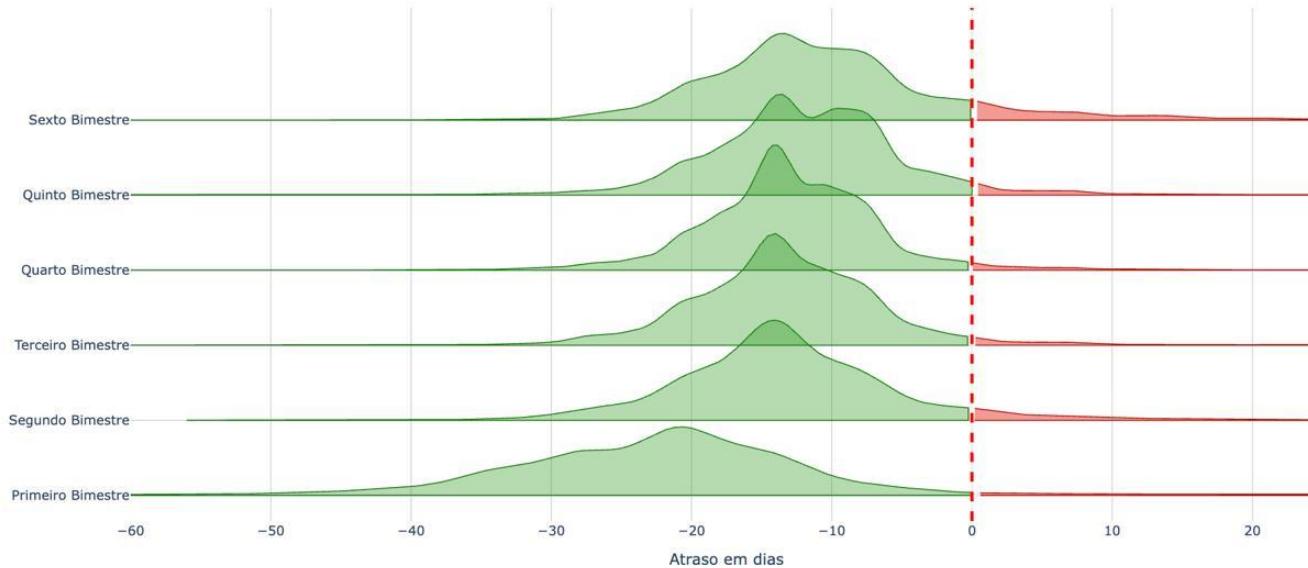
1. Qual o percentual de pedidos entregues atrasados por bimestre em 2017? Como é o comportamento semanal no bimestre de maior percentual?

bimestre_ano	dias_atraso	semana_ano	ano	numero_semana	media_dias_atraso		
14982	2017-B1	-18	0	2017-W01	2017	1	-25.050000
14995	2017-B1	-18	1	2017-W02	2017	2	-25.365079
14994	2017-B1	-18	2	2017-W03	2017	3	-29.289017
14993	2017-B1	-18	3	2017-W04	2017	4	-26.633333
14992	2017-B1	-18	4	2017-W05	2017	5	-24.997436

Consultas - Drill Down

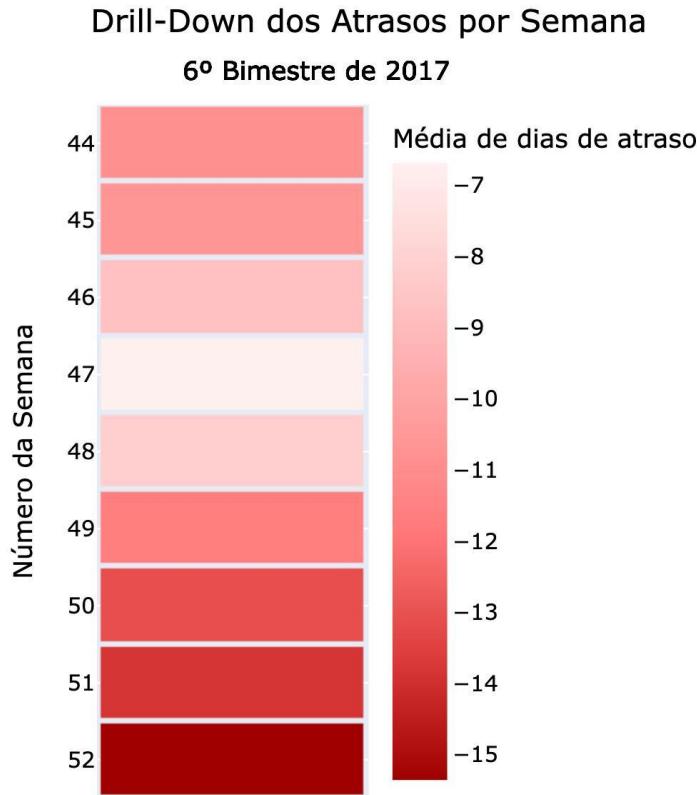
1. Qual o percentual de pedidos entregues atrasados por bimestre em 2017? Como é o comportamento semanal no bimestre de maior percentual?

Distribuição dos Tempos de Entrega em 2017 - em relação ao prazo previsto



Consultas - Drill Down

1. Qual o percentual de pedidos entregues atrasados por bimestre em 2017? Como é o comportamento semanal no bimestre de maior percentual?



Consultas - Roll Up + Slice/Dice

2. Como o churn rate (clientes que não voltaram a comprar após 3 meses) evoluiu (por mês) em 2018, por categoria de produto? Esse comportamento é semelhante ao observado no último ano (2017), por semestre?

ano	mes	categoria_produto	churn_insatisfeito	total_reviews	churn_percentual
2018	1	construction_tools_tools	2	2	100.0
2018	1	art	4	6	66.666666666666657
2018	1	books_imported	4	6	66.666666666666657
2018	1	market_place	4	8	50.0
2018	1	air_conditioning	4	8	50.0
2018	1	toys	96	216	44.444444444444443
2018	1	computers_accessories	136	350	38.857142857142854
2018	1	housewares	78	206	37.864077669902912
2018	1	furniture_mattress_and_upholstery	2	6	33.333333333333329
2018	1	food	4	12	33.333333333333329

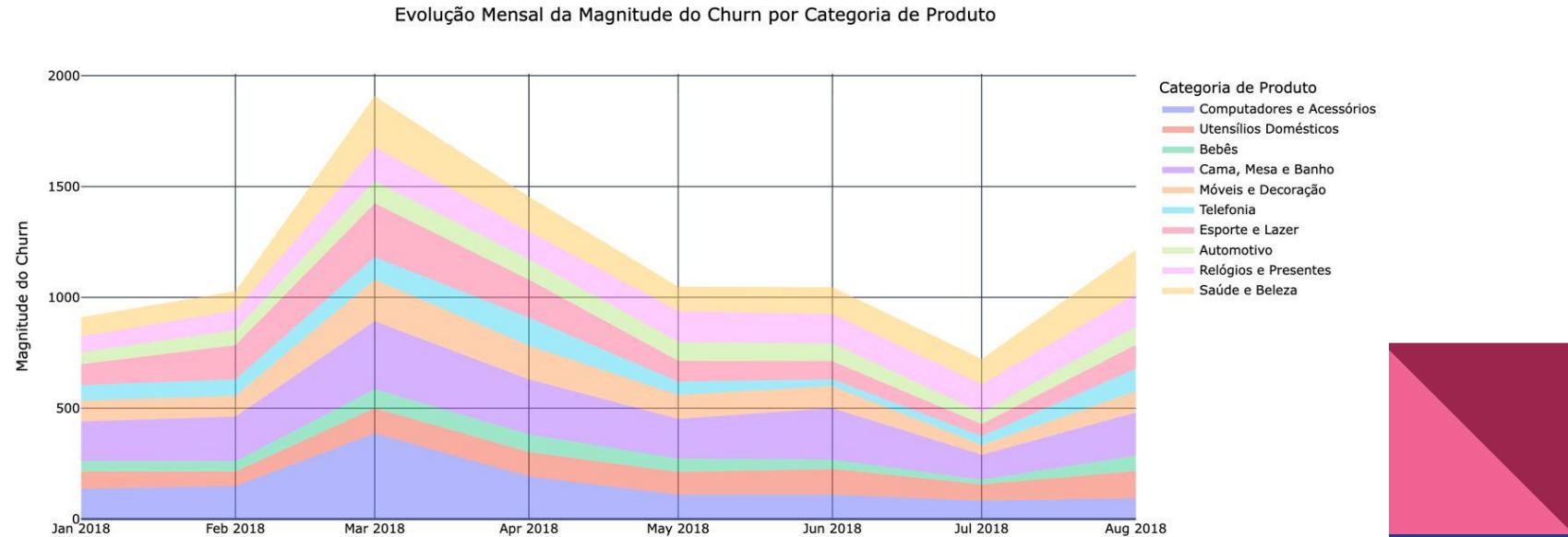
Consultas - Roll Up + Slice/Dice

2. Como o churn rate (clientes que não voltaram a comprar após 3 meses) evoluiu (por mês) em 2018, por categoria de produto? Esse comportamento é semelhante ao observado no último ano (2017), por semestre?

ano	numero_bimestre	bimestre_label	churn_rate_percentual
2017	1.0	B1	17.5561797752809
2017	2.0	B2	26.3184329482672
2017	3.0	B3	23.346828609986506
2017	4.0	B4	20.578683200472394
2017	5.0	B5	22.381756756756758
2017	6.0	B6	29.821531375935521
2018	1.0	B1	28.1591263650546
2018	2.0	B2	36.195002299555419
2018	3.0	B3	21.296009897927622
2018	4.0	B4	20.462197734925006

Consultas - Roll Up + Slice/Dice

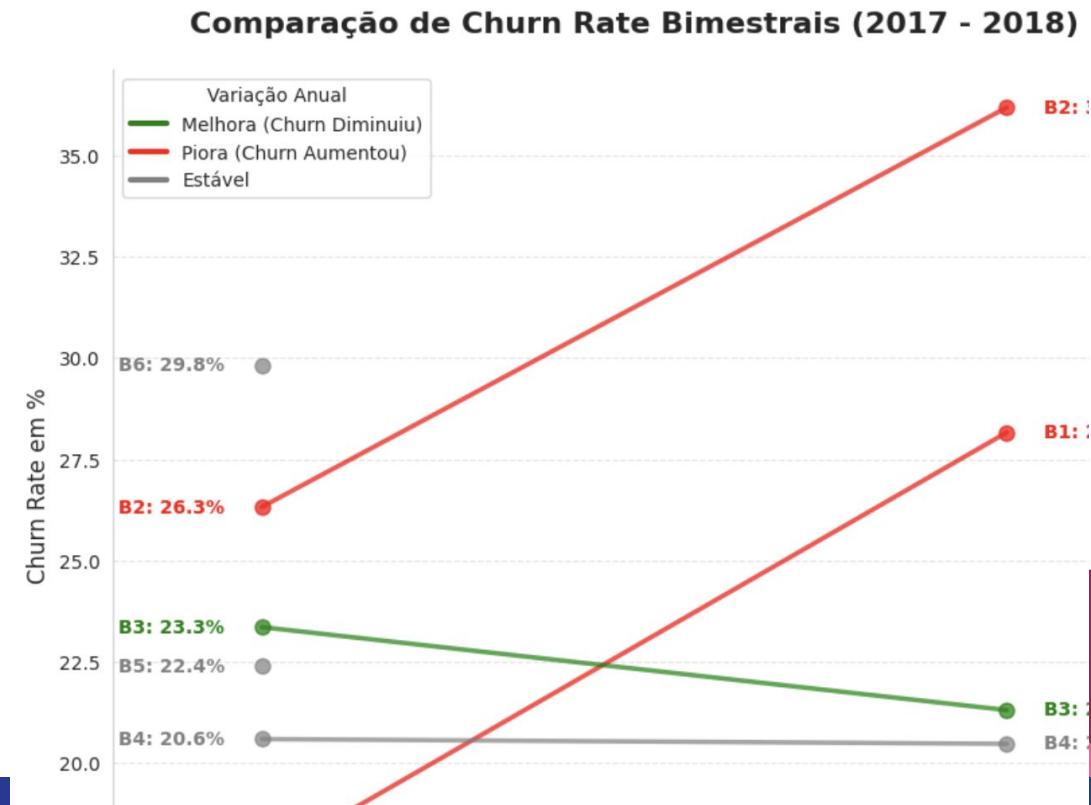
2. Como o churn rate (clientes que não voltaram a comprar após 3 meses) evoluiu (por mês) em 2018, por categoria de produto? Esse comportamento é semelhante ao observado no último ano (2017), por semestre?



Consultas - Roll Up + Slice/Dice

2. Como o churn rate (clientes que não voltaram a comprar após 3 meses) evoluiu (por mês) em 2018, por categoria de produto? Esse comportamento é semelhante ao observado no último ano (2017), por semestre?

Slice, Dice e Roll-Up!



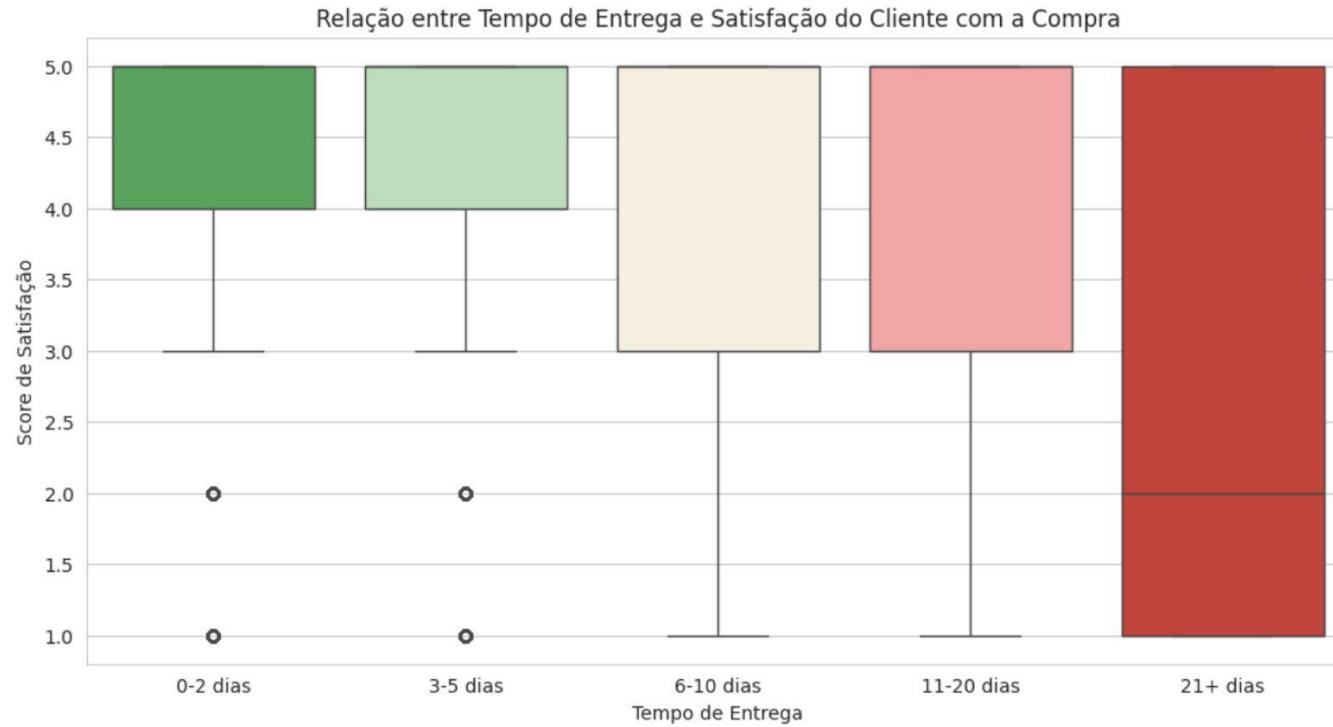
Consultas - Drill Across

3. Qual a relação entre tempo de entrega e satisfação geral do cliente com a compra?

order_id	satisfacao_score	cluster_satisfacao	tempo_entrega_dias
cef549d9-16f9-449b-a3e2-0654cc3be7ec	1	1	196
cef549d9-16f9-449b-a3e2-0654cc3be7ec	1	1	196
cef549d9-16f9-449b-a3e2-0654cc3be7ec	1	1	196
cef549d9-16f9-449b-a3e2-0654cc3be7ec	1	1	196
fada9add-c09c-4e94-8fe5-57c623ccd6de	1	1	195
fada9add-c09c-4e94-8fe5-57c623ccd6de	1	1	195
fada9add-c09c-4e94-8fe5-57c623ccd6de	1	1	195
fada9add-c09c-4e94-8fe5-57c623ccd6de	1	1	195
d20f46ad-8aa9-4c0e-a71a-ac17255ef8c7	1	1	191
d20f46ad-8aa9-4c0e-a71a-ac17255ef8c7	1	1	191

Consultas - Drill Across

3. Qual a relação entre tempo de entrega e satisfação geral do cliente com a compra?



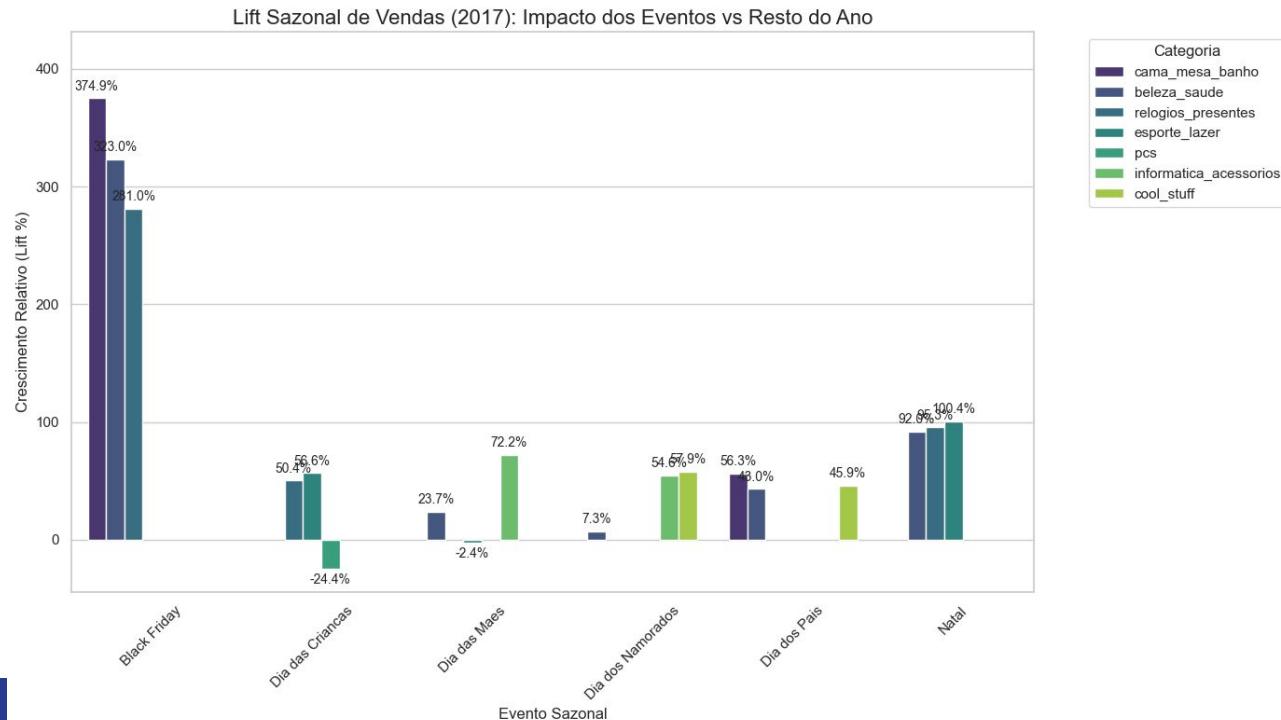
Consultas - Slice and Dice

4. Como é o comportamento das vendas em datas comemorativas em comparação ao resto do ano?

Evento Sazonal	Categoria	Volume/Vendas Evento	Média Diária Evento	Média diária Resto do ano	Crescimento Relativo
Black Friday	Cama/Mesa/Banho	R\$ 94.062,48	R\$ 11.757,81	R\$ 2.476,03	+374.9%
Dia das Crianças	Esporte/Lazer	R\$ 49.576,66	R\$ 3.305,11	R\$ 2.110,09	+56.6%
Dia das Mães	Informática/Acessórios	R\$ 46.312,26	R\$ 3.308,02	R\$ 1.920,89	+72.2%
Dia dos Namorados	Cool Stuff	R\$ 35.209,42	R\$ 3.200,86	R\$ 2.027,78	+57.9%
Dia dos Pais	Cama/Mesa/Banho	R\$ 57.190,72	R\$ 3.870,77	R\$ 2.476,03	+56.3%
Natal	Esporte/Lazer	R\$ 105.691,30	R\$ 4.227,65	R\$ 2.110,09	+100.4%

Consultas - Slice and Dice

4. Como é o comportamento das vendas em datas comemorativas em comparação ao resto do ano?



Slice e Dice!

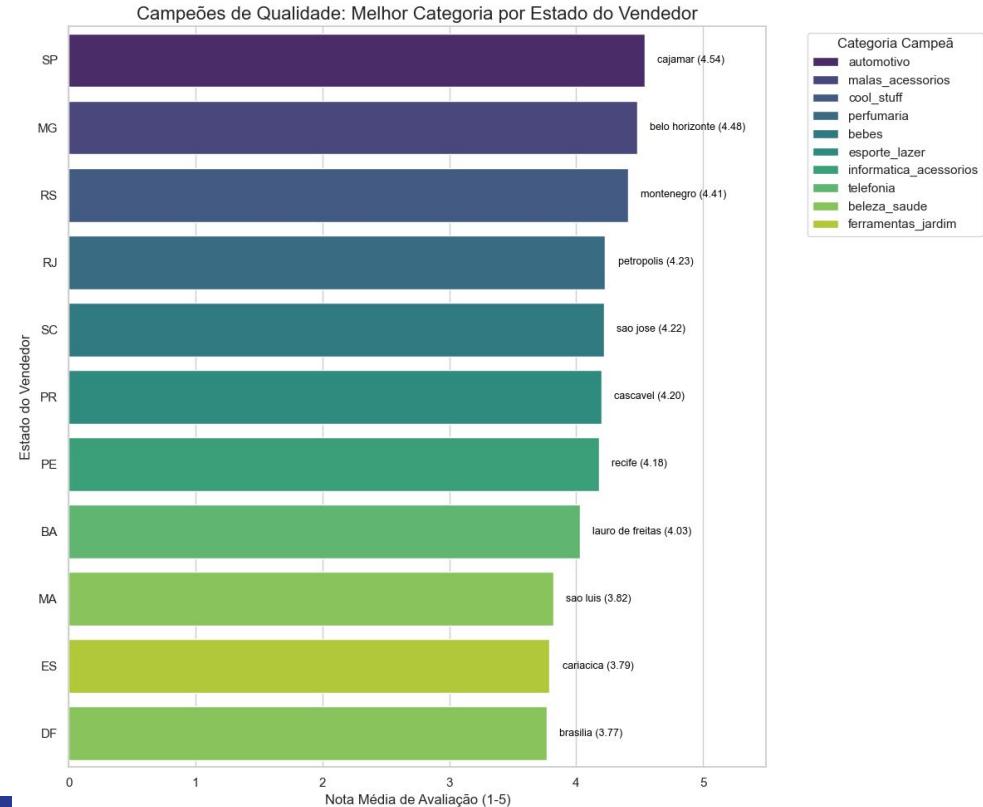
Consultas - Slice and Dice

5. Identificar o produto 'Campeão' de cada Estado baseado na avaliação (review).

Estado Vendedor	Cidade Vendedor	Categoria Melhor Avaliada	Qtd de Itens Vendidos	Receita Total	Nota Média Avaliação
SP	Cajamar	Automotivo	151	R\$ 18,755.80	4.54
MG	Belo Horizonte	Malas/Acessórios	127	R\$ 14,531.90	4.48
RS	Montenegro	Cool Stuff	130	R\$ 25,919.00	4.41
RJ	Petrópolis	Perfumaria	359	R\$ 70,909.30	4.23
SC	São José	Bebês	103	R\$ 6,866.10	4.22
PR	Cascavel	Esporte/Lazer	112	R\$ 16,016.49	4.20
PE	Recife	Informática/Acessórios	326	R\$ 53,897.00	4.18
BA	Lauro de Freitas	Telefonia	205	R\$ 54,617.68	04.03
MA	São Luís	Beleza/Saúde	395	R\$ 36,220.97	3.82
ES	Cariacica	Ferramentas jardim	173	R\$ 21,561.30	3.79
DF	Brasília	Beleza/Saúde	132	R\$ 14,899.07	3.77

Consultas - Slice and Dice

5. Identificar o produto 'Campeão' de cada Estado baseado na avaliação (review).



Slice e Dice!

Consultas - Pivot

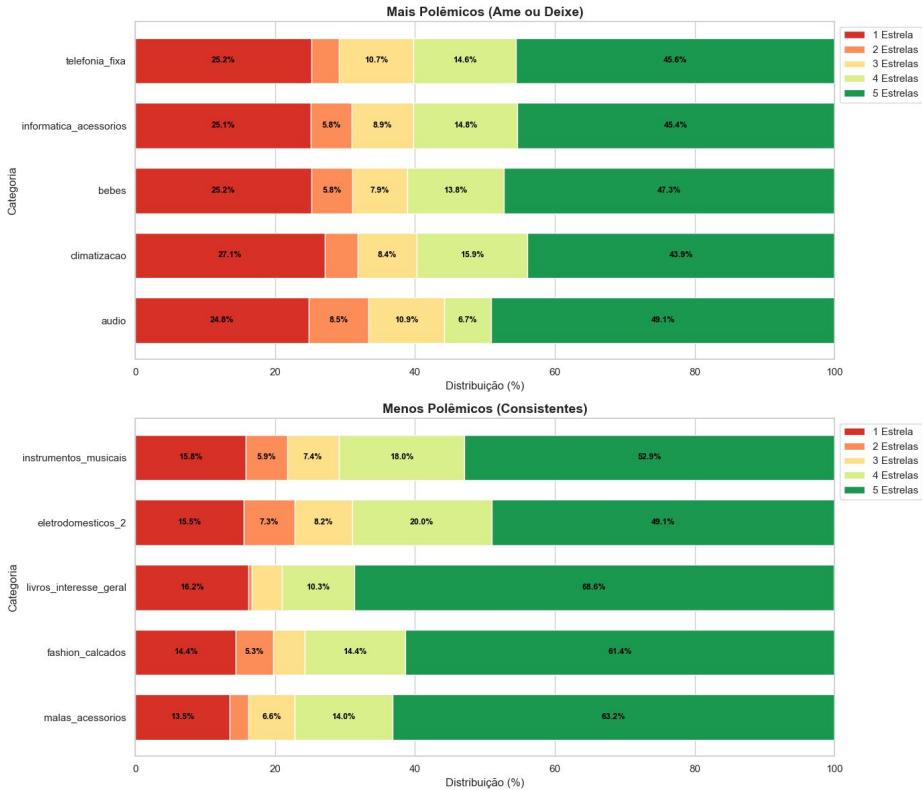
6. Quais são os produtos mais “polêmicos”, ou seja, aqueles que possuem maior desvio padrão entre suas avaliações ?

categoria	total_reviews	nota_media_real	desvio_padrao_polemica	share_5_estrelas	share_4_estrelas	share_3_estrelas	share_2_estrelas	share_1_estrela
audio	165	3.47	1.71	49.1%	6.7%	10.9%	8.5%	24.8%
climatizacao	107	3.45	1.69	43.9%	15.9%	8.4%	4.7%	27.1%
bebés	1134	3.52	1.68	47.3%	13.8%	7.9%	5.8%	25.2%
informatica_acesso								
rios	2967	3.49	1.67	45.4%	14.8%	8.9%	5.8%	25.1%
telefonia_fixa	103	3.51	1.67	45.6%	14.6%	10.7%	3.9%	25.2%

categoria	total_reviews	nota_media_real	desvio_padrao_polemica	share_5_estrelas	share_4_estrelas	share_3_estrelas	share_2_estrelas	share_1_estrela
malas_acessorios	437	4.11	1.42	63.2%	14.0%	6.6%	2.7%	13.5%
fashion_calcados	132	04.03	1.48	61.4%	14.4%	4.5%	5.3%	14.4%
livros_interesse_general	185	4.15	1.48	68.6%	10.3%	4.3%	0.5%	16.2%
eletrodomesticos_2	110	3.80	1.49	49.1%	20.0%	8.2%	7.3%	15.5%
instrumentos_musicais	272	3.86	1.50	52.9%	18.0%	7.4%	5.9%	15.8%

Consultas - Pivot

6. Quais são os produtos mais “polêmicos”, ou seja, aqueles que possuem maior desvio padrão entre suas avaliações ?



Pivot!

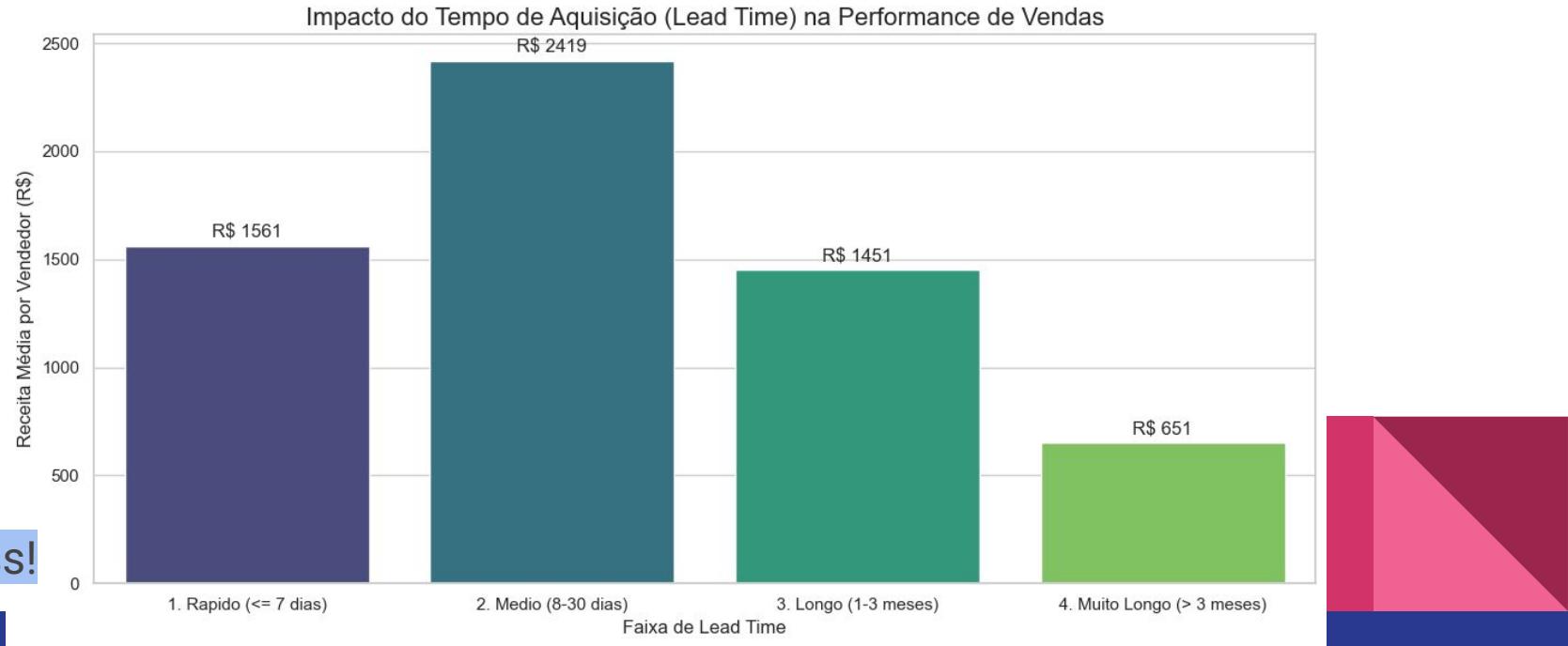
Consultas - Drill Across

7. Como o tempo até o fechamento do contrato com um novo vendedor impacta na receita dos fornecedores?

faixa_lead_time	qtd_vendedores	media_dias_fechamento	receita_total_grupo	receita_media_por_vendedor
1. Rapido (<= 7 dias)	160	3.1	R\$ 249,807.27	R\$ 1,561.30
2. Medio (8-30 dias)	137	15.4	R\$ 331,414.76	R\$ 2,419.09
3. Longo (1-3 meses)	52	52.9	R\$ 75,447.21	R\$ 1,450.91
4. Muito Longo (> 3 meses)	31	171.0	R\$ 20,182.24	R\$ 651.04

Consultas - Drill Across

7. Como o tempo até o fechamento do contrato com um novo vendedor impacta na receita dos fornecedores?



Conclusões

- Busca sistemática de ferramentas consistentes e fontes de dados;
- Definição dos processos de negócio que poderiam ser analisados a partir dos dados presentes e as métricas as quais seriam analisadas;
- Construção do modelo de como os dados estarão armazenados no SGBD (esquema estrela);
- Definição de quais questões seriam respondidas através das consultas estabelecidas.

Links Úteis

[Git com todo Código](#)

[Diagrama Estrela](#)