

Diretora: Jenifer Milena Pellin da Silva

Assessoras: Alessandra dos Santos Fernandes e Rafaela A. Schneider Hinsching

Assistente de Educação: Marisa R. Reichardt

Orientadora Educacional: Joice Daniela Santana Muraoka

Professor: Wendell Thomas Teske

Estudante: ARAM BOHMANN LEITE DA LUZ

RELATÓRIO DE GOVERNANÇA DE DADOS – PROJETO PETSHOP

1. INGESTÃO DE DADOS

O presente *database* apresenta na análise inicial: 16 colunas, 250.059 linhas registradas e quatro colunas categóricas sendo a forma de pagamento, categoria produto, estado e região. O tipo de dado mais presente é o texto, presente em 7 colunas, o segundo mais presente é o numérico com 6 colunas (excluído as colunas destinadas a data).

O tratamento de tais dados será requisitado, para correção de possíveis bugs futuros.

Dicionário de Dados:

16 Colunas • 250.059 Linhas

Projeto Pet Shop					
Nome original da Coluna	Nome atual da Coluna	Descrição	Tipo de Dados	Coluna Categórica	Valores Vazios
CÓDIGO PEDIDO	Código Pedido	Código do pedido	Numérico	Não	Não
região-país	Região	Região do Brasil	Texto	Sim	Não
produto	Produto	Produto	Texto	Não	Não
valor	Valor Unitário	Valor unitário do produto	Numérico	Não	Não
quantidade	Quantidade	Quantidade do produto	Numérico	Não	Não

valor_total_bruto	Valor Bruto	Valor total bruto do produto	Numérico	Não	Não
DIA	Dia	Dia da compra	Numérico	Não	Não
MES	Mês	Mês da compra	Numérico	Não	Não
ANO	Ano	Ano da compra	Numérico	Não	Não
estado	Estado	Estado brasileiro da compra	Texto	Sim	Não
formapagto	Forma de Pagamento	Forma de pagamento	Numérico	Sim	Não
centro_distribuicao	Centro de Distribuição	Centro de distribuição do produto	Texto	Não	Não
responsavelpedido	Responsável do Pedido	Responsável pelo pedido	Texto	Não	Não
valor_comissao	Valor da Comissão	Valor da comissão	Numérico	Não	Não
lucro_liquido	Lucro Líquido	Lucro líquido	Numérico	Não	Não
categoriaPROD	Categoria do Produto	Categoria do produto	Texto	Sim	Não

2. TRATAMENTO DE DADOS

A primeira etapa realizada, foi visualizar as possíveis inconsistências e erros da base. Assim, após a descrição do problema na análise, será demonstrado sua solução realizada.

O primeiro procedimento é no nome das colunas, como demonstrado na imagem abaixo, não há a nomeação correta, estando presente o nome correto da coluna na terceira linha da base.

Imagen 01 - Nomeação da coluna errada.

	A ^B _C Column1	A ^B _C Column2	A ^B _C Column3
1	Column1	Column2	Column3
2	'	223@	.
3	CÓDIGO PEDIDO	região-país	produto

Autor: Aram 2025.

Para a solução deste problema foram feitas 2 etapas, a primeira foi excluir as primeiras duas linhas, por não apresentar valor e consistência na base, e a segunda etapa, foi tornar a terceira linha no nome das colunas, como deveria ser originalmente.

Imagen 02 - Nomeação da coluna corrigida.

	1 ² ₃ CÓDIGO PEDIDO	A ^B _C região-país	A ^B _C produto
1	50284	Norte	Biscoito True Champion 300g
2	50285	Norte	Biscoito True Champion 300g
3	50286	Norte	Biscoito True Champion 300g

Autor: Aram 2025.

O segundo procedimento encontrado é na padronização dos nomes das colunas, cada coluna possui uma forma de ser escrita, com espaço, com traço ou até com underline. Que pode causar erros de entendimento. Como demonstra a imagem abaixo:

Imagen 03 - Padronização da coluna errada.

A ^B _C região-país	A ^B _C formapagto	A ^B _C centro_distribuicao
---	--	---

Autor: Aram 2025.

Para a solução deste problema, os nomes das colunas foram padronizados da seguinte forma: “Nome Sobrenome”. Como demonstra a imagem abaixo:

Imagen 04 - Padronização da coluna corrigida.

A ^B _C Estado	A ^B _C Forma de Pagamento	A ^B _C Centro de Distribuição
------------------------------------	--	--

Autor: Aram 2025.

O terceiro procedimento encontrado foi na padronização dos nomes dos elementos das colunas. Por exemplo, na coluna Região apresenta valores “centrooeste” e valores “Centro-Oeste”, representando a mesma coisa e dois elementos distintos, como demonstra alguns casos abaixo:

Imagen 05 - Exemplos de padronização dos valores das colunas erradas.

Norte	
Nordeste	
Sudeste	
Centro Oeste	
Sul	
centrooeste	
noRtE	
Centro-Oeste	

2020	
dois mil e vinte	

Rio Grande do Sul	
Piauí	
Ceará	
amapá	
Amapá	
ceará	

Autor: Aram 2025.

Para a solução deste problema, houve a devida correção e padronização de todo os valores da base, como demonstrado na imagem abaixo:

Imagen 06 - Exemplos de padronização dos valores das colunas corrigidas.

Norte	
Nordeste	
Sudeste	
Centro Oeste	
Sul	

2020	
------	--

Rio de Janeiro	
Espírito Santo	
Paraná	
Santa Catarina	
Rio Grande do Sul	
Piauí	

Autor: Aram 2025.

O quarto procedimento foi nas colunas com números decimais. Eles apresentam seus pontos flutuante com “.”, e não “,” que é o padrão brasileiro, além

de ter seu tipo de dados em string, como apresenta a imagem abaixo:

Imagen 07 - Exemplos de colunas com ponto flutuante e tipo de dados erradas.

A ^B _C	Valor	A ^B _C	Valor Total Bruto
	24.64		49.28
	23.52		47.04
	24.64		98.56
	21.28		85.12

A ^B _C	Valor da Comissão	A ^B _C	Lucro Líquido
	1.76		26.4
	1.68		25.2
	3.52		52.8

Autor: Aram 2025.

Para a solução deste problema, houve a troca do ponto para a vírgula, e a transformação do tipo de dados, como mostra a imagem abaixo:

Imagen 08 - Exemplos de colunas com ponto flutuante e tipo de dados corrigidos.

1.2	Valor	1.2	Valor Total Bruto
	24,64		49,28
	23,52		47,04
	24,64		98,56

1.2	Valor da Comissão	1.2	Lucro Líquido
	1,76		26,4
	1,68		25,2
	3,52		52,8

Autor: Aram 2025.

O quinto procedimento foi nas colunas destinadas a datas, que estavam separadas e com o tipo de dados errados, que pode causar futuros erros nas criações de gráficos, como demonstra a imagem abaixo. E além disso, havia datas irreais, como “31/02”, que devia ser tratado para não haver erros futuros ao transformar o tipo de dados, entretanto, foi percebido datas que só existem em anos bissexto, como o “29/12”.

Imagen 09 - Colunas com data errado.

A ^B _C Dia	A ^B _C Mês	A ^B _C Ano
12	03	2020
10	04	2020
07	08	2020

Autor: Aram 2025.

Para a solução deste problema, foi mesclado as três colunas presentes, e separados por “/”, e posteriormente trocado o tipo de dado para *date*, como demonstra a imagem abaixo. Entretanto, antes da troca do tipo de dado, foi convertido as datas irreais apresentadas para a data mais próxima.

Imagen 10 - Colunas com data corrigida.

Data
12/03/2020
10/04/2020
07/08/2020
10/01/2020

Autor: Aram 2025.

O sexto procedimento foi nas colunas numéricas que possuíam valores negativos. Algo que não possuía consistência na base, como demonstra a imagem abaixo:

Imagen 11 - Colunas com números negativos.

42	-44	-1830
42	-900	-37440

Autor: Aram 2025.

Para a solução deste problema, o tipo de dado foi transformado em texto, para fazer a remoção do “-”, e posteriormente voltado ao decimal.

2.1. ESTRATÉGIA(S) PARA RESOLUÇÃO DE OUTLIERS

Os dados *outliers* não apareceram em todas as colunas, entretanto, se fez presente na coluna valor bruto, como um dado ter registrado 879.789 mil reais em uma única venda, um valor que deveria ser apenas 117 reais. Para a solução do problema, foi realizado o mesmo comando dax futuramente apresentado para

solucionar os dados *missing* da mesma coluna. Onde o valor bruto é o valor unitário vezes a quantidade, como demonstra a figura abaixo:

Imagen 12 - Colunas com números negativos.

Valor Unitário	Quantidade	Valor Total Bruto	Valor Total Bruto Corrigido
39	3	879789	117
26	4	876867	104

Autor: Aram 2025.

2.2. ESTRATÉGIA(S) PARA RESOLUÇÃO DE MISSINGS

A base atual apresenta uma coluna com dados *missing*, a quantidade do produto, essa coluna é sensível a alterações, pois afeta diretamente o desempenho das vendas, como demonstra a imagem abaixo:

Imagen 13 - Coluna com *missing*.

A ^B _C Quantidade
4
3
2

Autor: Aram 2025.

Para a solução deste problema, foi pensando em uma estratégia peculiar. A base apresenta para nós o produto, seu valor, valor bruto e a quantidade, é possível relacionar o valor bruto do produto com a quantidade e o valor unitário. Ou seja, se em um registro apresentou 4 quantidades com valor unitário de 10 reais e um valor bruto de 40 reais, e abaixo um dado *missing* de quantidade com um valor também de 10 reais do valor bruto de 40 reais, é possível deduzir que a quantidade se manterá semelhante. Como demonstra o exemplo abaixo:

Imagen 14 - Coluna com *missing* exemplo da lógica.

1.2 Valor Unitário	1 ² 3 Quantidade	1.2 Valor Total Bruto
22,4	null	44,8
22,4	2	44,8

Autor: Aram 2025.

Pensando nessa lógica, foi feita a devida análise e adição dos valores através de comandos dax, a quantidade é o valor total bruto dividido pelo valor unitário, pensando sempre em manter o padrão dos outros registros na base, como demonstra a imagem abaixo. E além disso, foi tratado para não haver quantidades decimais, algo sem consistência, que foi solucionado arredondando o valor.

Imagen 15 - Dax da coluna.

```
1 Quantidade Corrigido =  
2 IF('base_vendas-petshop (1)'[Valor Total Bruto] > 0,  
3     DIVIDE('base_vendas-petshop (1)'[Valor Total Bruto Corrigido], 'base_vendas-petshop (1)'[Valor Unitário]),  
4     'base_vendas-petshop (1)'[Quantidade]  
5 )
```

Autor: Aram 2025.

Imagen 15 - Coluna com *missing* corrigida.

1.2 Valor Unitário	1.2 Quantidade	1.2 Valor Total Bruto
22,4	2	44,8
22,4	2	44,8

Autor: Aram 2025.

A segunda coluna que apresenta dados *missing* é o valor bruto, e pensando na mesma lógica anterior, o valor missing do valor bruto será o valor unitário vezes a quantidade do produto. Para aplicar essa lógica, foi feito o dax do mesmo:

Imagen 16 - Dax da coluna.

```
1 Valor Total Bruto Corrigido = 'base_vendas-petshop (1)'[Valor Unitário] * 'base_vendas-petshop (1)'[Quantidade]
```

Autor: Aram 2025.

Entretanto, mesmo após todas as operações para corrigir os dados missing, havia casos isolados que a quantidade original era 0, e o valor bruto original também é 0, não sendo possível salvar tais valores usando a lógica anterior. E além disso, colunas como valor da comissão, possui valores 0, mas não foi considerado como *missing* por fazer sentido e consistência na base. Um adendo que a base possuía 20 linhas com erros na quantidade, e eles foram alterados para se tornarem *missing* e tratados da mesma forma usando o comando dax anterior, que solucionou o problema.

3. MANIPULAÇÃO DE DADOS

Para iniciarmos o enriquecimento de dados, foi realizado indicadores kpi's, medidas que comportava a quantidade total vendido, o valor brutol total, o lucro líquido total e a margem de lucro (%), como demonstra as figuras abaixo:

Imagen 17 - Medidas de somas.

```
1 Quantidade Total Vendido = SUM('base_vendas-petshop (1)'[Quantidade Corrigido])
```

```
1 Valor Bruto Total = SUM('base_vendas-petshop (1)'[Valor Bruto Corrigido])
```

```
1 Lucro Líquido Total = SUM('base_vendas-petshop (1)'[Lucro Líquido])
```

```
1 Margem de Lucro (%) = (SUM('base_vendas-petshop'[Lucro Líquido]) / SUM('base_vendas-petshop'[Valor Bruto Corrigido])) * 100
```

Autor: Aram 2025.

Posterior a isso, foi criado uma tabela para trabalhar com datas, nomeado como dCalendario, relacionando com a data da tabela principal. Como demonstra nas imagens abaixo:

Imagen 18 - Dax para a criação da tabela Calendário.

```
1 dCalendar =  
2 ADDCOLUMNS(  
3     CALENDAR(MIN('base_vendas-petshop (1)'[Data]), MAX('base_vendas-petshop (1)'[Data])),  
4     "Ano", YEAR([Date]),  
5     "Mês", MONTH([Date]),  
6     "Dia Semana", WEEKDAY([Date]),  
7     "Dia", DAY([Date]),  
8     "Mês Extenso", FORMAT([Date],  
9     "MMMM"), "Mês Abreviado", FORMAT([Date], "MMM"),  
10    "Dia Extenso", FORMAT([Date], "DDDD"),  
11    "Dia Abreviado", FORMAT([Date], "DDD"),  
12    "Trimestre", QUARTER([Date])  
13 )
```

Autor: Aram 2025.

Imagen 19 - relacionamento das tabelas.

Novo relacionamento

Selecione tabelas e colunas relacionadas umas às outras.

Da tabela

dCalendario

Ano	Date	Dia	Dia Abreviado	Dia Extenso	Dia Semana	Mês
2020	01/07/2020 0...	1	qua	quarta-feira	4	7
2020	02/07/2020 0...	2	qui	quinta-feira	5	7
2020	03/07/2020 0...	3	sex	sexta-feira	6	7

Para a tabela

base_vendas-petshop (1)

Categoria do ...	Centro de Dis...	Código Pedido	Data	Estado	Forma de Pag...	Lucro Líqu...
Acessório	Gold Beach	50741	sexta-feira, 17...	Ceará	Dinheiro	51
Acessório	Gold Beach	53141	quarta-feira, 2...	Ceará	Dinheiro	49
Acessório	Gold Beach	55541	terça-feira, 17...	Ceará	Dinheiro	51

Cardinalidade

Um para muitos (1:*)

Direção do filtro cruzado

Único

Autor: Aram 2025.

Após isto, foi corrigido às regiões dos estados criando uma nova coluna utilizando o *SWITCH*, abaixo estão dois exemplos de dax possíveis para a criação desta coluna:

Imagen 20 - Parte do dax usado.

```

1 Região Corrigida =
2 SWITCH (
3   TRUE,
4   'base_vendas-petshop (1)'[Estado] = "Acre", "Norte",
5   'base_vendas-petshop (1)'[Estado] = "Amapá", "Norte",
6   'base_vendas-petshop (1)'[Estado] = "Amazonas", "Norte",
7   'base_vendas-petshop (1)'[Estado] = "Pará", "Norte",
8   'base_vendas-petshop (1)'[Estado] = "Rondônia", "Norte",
9   'base_vendas-petshop (1)'[Estado] = "Roraima", "Norte",
10  'base_vendas-petshop (1)'[Estado] = "Tocantins", "Norte",
11  'base_vendas-petshop (1)'[Estado] = "Alagoas", "Nordeste",
12  'base_vendas-petshop (1)'[Estado] = "Bahia", "Nordeste",
13  'base_vendas-petshop (1)'[Estado] = "Ceará", "Nordeste",
14  'base_vendas-petshop (1)'[Estado] = "Maranhão", "Nordeste",
15  'base_vendas-petshop (1)'[Estado] = "Paraíba", "Nordeste",
16  'base_vendas-petshop (1)'[Estado] = "Piauí", "Nordeste",
17  'base_vendas-petshop (1)'[Estado] = "Pernambuco", "Nordeste",

```

```

1 Região_Corrigida =
2 SWITCH (
3   TRUE,
4   --Norte
5   'base_vendas-petshop'[Estado] = "Acre" || 'base_vendas-petshop'[Estado] = "Amapá" || 'base_vendas-petshop'[Estado] =
6   "Amazonas" || 'base_vendas-petshop'[Estado] = "Pará" || 'base_vendas-petshop'[Estado] = "Rondônia" || 'base_vendas-petshop'
7   [Estado] = "Roraima" || 'base_vendas-petshop'[Estado] = "Tocantins", "Norte",
8   -- Nordeste
9   'base_vendas-petshop'[Estado] = "Alagoas" || 'base_vendas-petshop'[Estado] = "Bahia" || 'base_vendas-petshop'[Estado] =
10  "Ceará" || 'base_vendas-petshop'[Estado] = "Maranhão" || 'base_vendas-petshop'[Estado] = "Paraíba" || 'base_vendas-petshop'
11  [Estado] = "Piauí" || 'base_vendas-petshop'[Estado] = "Pernambuco" || 'base_vendas-petshop'[Estado] = "Rio Grande do
12  Norte" || 'base_vendas-petshop'[Estado] = "Sergipe", "Nordeste",
13  -- Centro Oeste
14  'base_vendas-petshop'[Estado] = "Distrito Federal" || 'base_vendas-petshop'[Estado] = "Goiás" || 'base_vendas-petshop'
15  [Estado] = "Mato Grosso" || 'base_vendas-petshop'[Estado] = "Mato Grosso do Sul", "Centro Oeste",
16  -- Sudeste
17  'base_vendas-petshop'[Estado] = "Espírito Santo" || 'base_vendas-petshop'[Estado] = "Minas Gerais" || 'base_vendas-petshop'
18  [Estado] = "Rio de Janeiro" || 'base_vendas-petshop'[Estado] = "São Paulo", "Sudeste",
19  -- SUL
20  'base_vendas-petshop'[Estado] = "Paraná" || 'base_vendas-petshop'[Estado] = "Rio Grande do Sul" || 'base_vendas-petshop'
21  [Estado] = "Santa Catarina", "Sul"
22 )

```

Autor: Aram 2025

Por fim, em busca de uma interação maior no dashboard, foi realizado um título que muda de acordo com o ano selecionado, criado com o seguinte comando dax:

Imagen 20 - Parte do dax usado.

```

1 TítuloDashboard =
2 "Vendas Pet Shop - " &
3 SELECTEDVALUE(dCalendario[Ano], "Todos os Anos")

```

Autor: Aram 2025

4. ANÁLISE DE DADOS & TOMADA DE DECISÃO

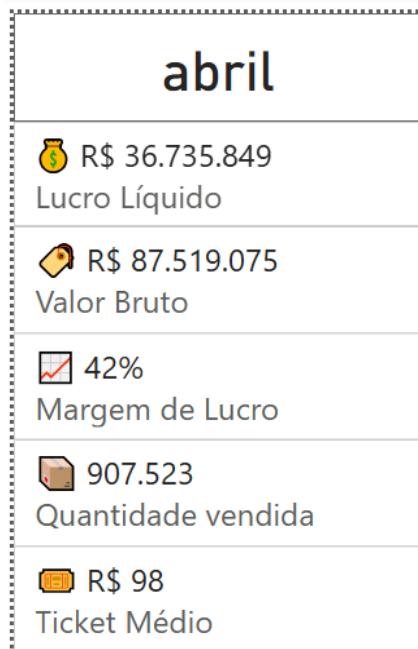
4.1. CONSTRUÇÃO DO DASHBOARD

Iniciamos a criação do dashboard sobre o “Desempenho de Vendas PetShop”, planejando os visuais que serão utilizados, sendo eles:

- Indicadores KPI's: Lucro total - Ticket Médio - Quantidade vendida - Margem de Lucro (%).
- Gráfico de Linhas: Será utilizado para a tendência das vendas por mês.
- Gráfico de barras: Distribuidora que mais vendeu produtos - Categoria que mais vendeu produtos - Top Produtos.
- Filtros: Categoria - Ano - Mês - Região - Estado.
- Mapa coroplético: Será utilizado para avaliar as diferentes performance das regiões e estados

Visando uma maior capacidade de análise, será criado cartões de dicas de ferramenta que apresentará dados mais detalhados, como demonstra a imagem abaixo:

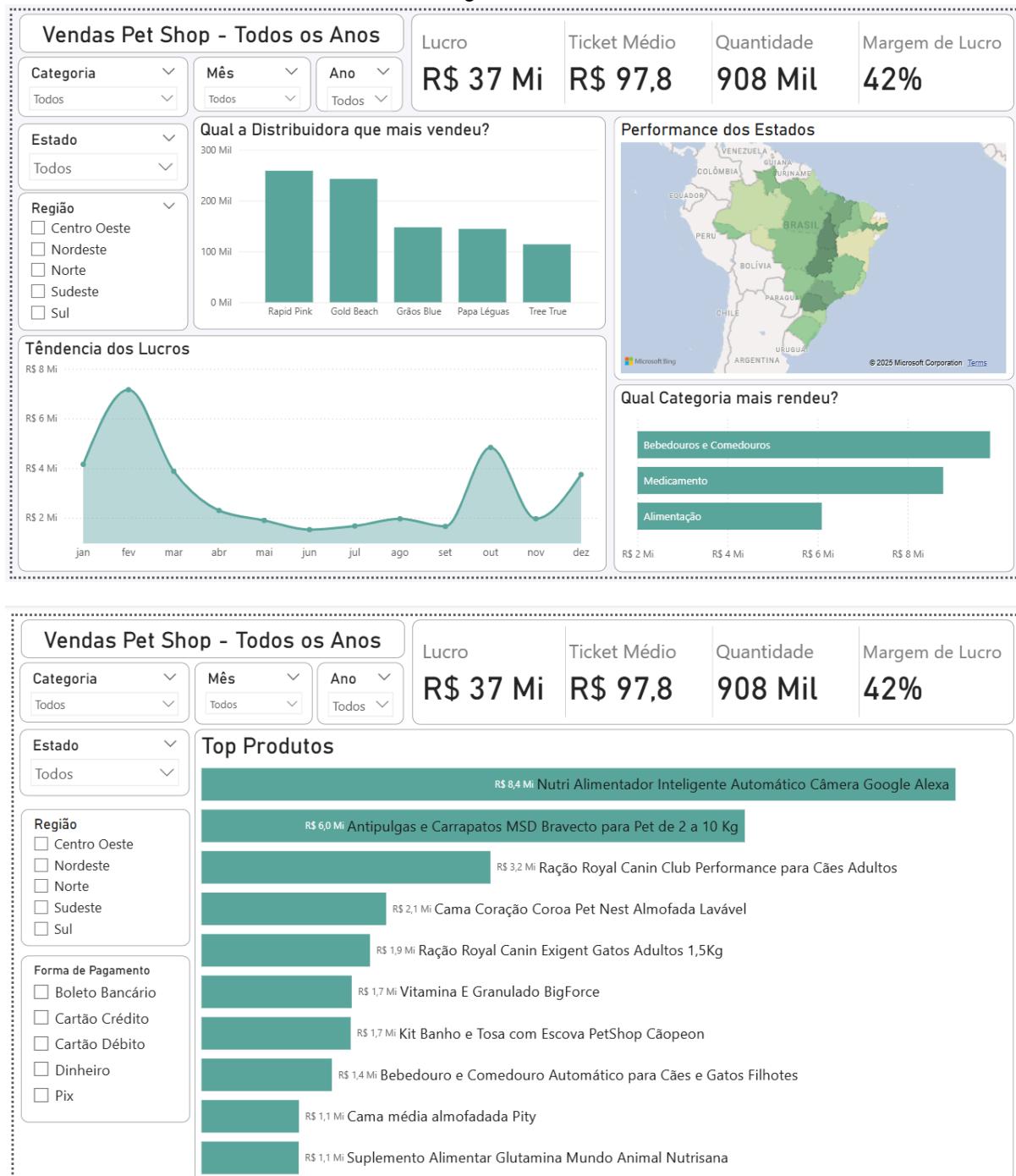
Imagen 21 - Exemplo da Dica de Ferramenta.



Autor: Aram 2025

Agora com os planejamentos visuais completos, foi realizado de fato o dashboard, sendo um para análise gerais, e outro específico para analisar produtos. Como apresentado nas imagens abaixo:

Imagen 22 - Dashboard.



Autor: Aram 2025

4.2. Considerações Finais e Tomada de Decisão

Com base em uma análise do dashboard apresentado anteriormente, é possível identificar tendências, oportunidades e estratégias de melhoria para o desempenho geral das vendas do Pet Shop, a seguir, esta algumas tomadas de decisões identificadas:

- **Desempenho Regional:**

O mapa de performance dos estados evidencia que algumas regiões possuem baixo volume de vendas em comparação a outras, e que, estados como Sergipe e Rio Grande do Norte não possuem clientes ativos dos produtos.

Ações recomendadas para a solução desses problemas são: Intensificar campanhas de marketing e parcerias locais nessas áreas, ampliando os distribuidores, oferecendo promoções regionais e eventos de abertura nos estados ainda sem cliente e identificar alterações possíveis no produto para se encaixar nas necessidades de cada estado.

- **Centro de Distribuição:**

As distribuidoras Rapid Pink e Gold Beach destacam-se em volume de vendas, enquanto outras como Tree True apresentam desempenho menor.

Ações recomendadas para a solução desses problemas são: Revisar contratos e condições comerciais das distribuidoras de menor performance, buscando negociar melhores prazos, descontos ou exclusividades de produtos e a geração de uma maior receita.

- **Produtos:**

Os produtos das categorias Bebedouros e Comedouros são os que mais geram receita, seguidos por Medicamentos e Alimentação.

Ações recomendadas para a ampliação das vendas: Aumentar o mix de produtos e estoques das categorias mais rentáveis e realizar campanhas de cross-selling.

- **Tendência Temporal:**

A análise mensal dos lucros mostra picos em fevereiro e outubro, com quedas em meses intermediários.

Ações recomendadas para a ampliação das vendas: Investir em promoções sazonais nos meses de baixa (junho a agosto), como programas de fidelidade, descontos progressivos e campanhas temáticas (inverno pet, mês do banho e tosa etc...).

Em suma ao projeto, houve em êxito a capacidade de tratamento e enriquecimento de dados até chegarmos nas análises apresentadas anteriormente, que nos ajudaram a responder diversas perguntas sobre o desempenho das vendas e possíveis tomadas de decisões.