

---

2025

# RELATÓRIO

## Analise de Dados

CRIADO POR

**Lucas Ferreira**

linkedin : lfaferreira

# Índice

## [1 Introdução](#)

[1.1 Objetivo do relatório](#)

[1.2 Carga Incremental](#)

[1.3 Infraestrutura](#)

## [2. Decisões Técnicas Adotadas](#)

[2.1 Estratégia de Transformação dos dados](#)

[2.2 Estrutura Simples e Linear](#)

[2.3 Reutilização de Funções](#)

## [3. Descrição das Transformações Aplicadas](#)

[3.1 Tipagem e Conversões de Dados](#)

[3.2 Limpeza de Dados](#)

## [4. Análise Descritiva dos Dados](#)

[4.1 Dados Numericos](#)

[4.2 Dados Categoricos](#)

[4.3 Dados Temporais](#)

## [5 Hipóteses e Padrões Identificados](#)

[5.1 Hipótese 1](#)

[5.2 Hipótese 2](#)

[5.3 Hipótese 3](#)

## [6. Observações Gerais e Insights](#)

[6.1 Principais Recomendações](#)

## [7. Conclusão](#)

# 1. Introdução

No cenário comercial atual, a capacidade de tomar decisões estratégicas baseadas em dados é um diferencial competitivo. Este relatório tem como finalidade apoiar a liderança da área comercial na análise do desempenho individual e coletivo do time de vendas nos últimos meses, com foco em identificar padrões, oportunidades de melhoria e alavancar o crescimento sustentável.

Os dados analisados incluem métricas críticas como volume de vendas, MRR (Monthly Recurring Revenue) e comissões pagas, seguindo a régua de remuneração variável estabelecida (1% para até 10 vendas e 1,5% a partir de 30 vendas). A análise foi estruturada em duas etapas:

- Organização dos dados no Google Sheets, com detalhamento mensal e por vendedor.
- Visualização interativa no Looker Studio, permitindo filtros dinâmicos e comparações temporais.

O relatório não apenas resume o desempenho passado, mas também propõe ações práticas para otimizar resultados, garantindo que a estratégia comercial esteja alinhada às metas de crescimento e à motivação da equipe.

## 1.1 Objetivo do relatório

### 1. Primário:

- a. Consolidar e organizar dados de desempenho comercial de forma clara e acessível, destacando métricas-chave por vendedor e por mês.
- b. Identificar padrões de performance, como sazonalidade, correlação entre vendas e MRR e comportamentos fora do esperado.

### 2. Secundário:

- a. Oferecer insights sobre a eficácia do modelo de comissão atual
- b. Recomendar ajustes operacionais e estratégicos para a liderança, como treinamentos direcionados ou revisão de metas.

### 3. Técnico:

- a. Demonstrar a relação entre a remuneração variável e os resultados gerados
- b. Facilitar a tomada de decisão com visualizações interativas e dashboards atualizáveis.

## 1.2 Contexto dos dados analisados

Este relatório analisa dados de vendas registrados entre março de 2022 e agosto de 2023, provenientes de uma planilha no Google Sheets, abrangendo negócios fechados identificados por `id` e `pipeline_id`. A amostra inclui métricas financeiras como MRR (*Monthly Recurring Revenue*), que representa a receita mensal recorrente dos contratos, e NRR (*Non-Recurring Revenue*), referente a receitas únicas como taxas de implementação. Também estão incluídos dados de churn rate, definidos por contrato — uma modelagem incomum, já que tradicionalmente o churn é calculado no nível do cliente, o que sugere atenção para validação metodológica.

As métricas-chave consideradas são: volume total de vendas (contagem de `id`), soma total de MRR e NRR, churn rate médio por contrato, além da distribuição de tipos de planos vendidos (como Astrea Starter, Astrea Pro, Smart) e periodicidade (predominantemente mensal).

Foram assumidas premissas importantes: cada `id` representa uma venda única; as comissões de vendedores são calculadas exclusivamente sobre o MRR, seguindo a regra de 1% para até 10 vendas mensais e 1,5% para 30 vendas ou mais; e a periodicidade mensal significa que o MRR reflete a receita gerada a cada mês (por exemplo, um plano de R\$ 669/mês).

Durante o preparo dos dados, foram aplicadas transformações essenciais: agregação de vendas e receitas por vendedor e mês, cálculo das comissões conforme a regra estipulada, e revisão do churn rate para garantir que ele não esteja sendo duplicado entre contratos pertencentes ao mesmo cliente.

### 1.3 Metodologia

O trabalho foi desenvolvido utilizando Google Sheets para o armazenamento inicial dos dados brutos, Jupyter Notebook (via VS Code) como ambiente principal para transformações, análises e engenharia de features (com Python e bibliotecas como pandas, numpy e matplotlib), e Looker Studio para a criação de dashboards interativos.

O fluxo de processamento seguiu as etapas: primeiro, os dados brutos (armazenados em `data/raw/raw_data.csv`) — contendo colunas como `id`, `won_time`, `vendedor`, `mrr`, `nrr`, entre outras — foram carregados no notebook `note_01_data_transform.ipynb`. Lá, aplicaram-se limpezas básicas (como tratamento de datas e remoção de duplicatas), criação de variáveis temporais (ano, mês, dia da semana), e análises estatísticas descritivas para entender distribuições e identificar outliers, como churn rates acima de 0,3.

Na etapa de feature engineering, foram calculados KPIs relevantes (como receita total =  $MRR + NRR$ , e proporção  $MRR/NRR$ ), categorizados desempenhos (ex.: classificação de vendas em `abaixo_da_meta` ou `meta_atingida`) e derivadas métricas temporais, como dias sem vendas e tempo médio entre vendas. As saídas processadas foram salvas em `data/processed/processed_data.csv` para análise exploratória, enquanto o arquivo final `analise_vendas.xlsx` (em `data/delivery/`) foi preparado para ser consumido pelo Looker Studio, já agregado por mês e vendedor.

As técnicas estatísticas incluíram agregações mensais (vendas totais, MRR, NRR), cálculo de médias móveis para métricas-chave como churn rate e vendas mensais, além de segmentações usando quartis e análise temporal para identificar padrões de sazonalidade e produtividade.

Por fim, o Looker Studio foi integrado diretamente ao arquivo `processed_data.csv` para garantir boa performance, gerando visualizações como gráficos de tendência mensal (ex.: MRR total ao longo do tempo), tabelas dinâmicas filtráveis por vendedor, receita total, etc para explorar correlações entre métricas, como volume de vendas versus churn rate.

## 2. Decisões Técnicas Adotadas

### 2.1 Estratégia de Transformação dos dados

Após uma análise detalhada das necessidades do projeto e das restrições orçamentárias, foi definida e implementada uma estratégia baseada no uso de ferramentas Open Source, como Python, VS Code, Jupyter Notebooks, Looker Studio e Google Sheets.

1. **Otimização de Custos:** A escolha por soluções Open Source permitiu reduzir significativamente os custos com licenciamento e assinaturas, direcionando os recursos disponíveis para atividades essenciais de análise e entrega de valor..
2. **Eficiência Operacional:** A adoção dessas ferramentas também proporcionou ganhos operacionais, graças ao suporte de comunidades ativas que continuamente contribuem para a evolução, documentação e resolução de problemas relacionados a essas tecnologias.

### 2.2 Estrutura Simples e Linear

O arquivo notebook foi construído com uma estrutura simples e linear, visando facilitar tanto a compreensão do código quanto a manutenção da pipeline. Ele está dividido em setores bem definidos:

- **Importação e Configuração Inicial:**  
Abrange os pacotes e parâmetros necessários para o funcionamento do pipeline.
- **Pré-processamento:**  
Inclui etapas como verificação dos tipos de dados, identificação e tratamento de valores nulos (`check NA`, `fill NA`), filtragem de linhas, ajuste de nomes de colunas e padronização dos tipos de dados. Aqui também são feitas as primeiras identificações de possíveis incongruências nos dados.
- **Análise Estatística Descritiva:**  
Os dados são separados em numéricos, categóricos e temporais, permitindo visualizar suas distribuições e entender melhor seus comportamentos, direcionando as decisões de tratamento e transformação.
- **Criação de Novas Features:**  
Nesta etapa, são geradas novas variáveis que enriquecem o conjunto de dados, agregando informações complementares relevantes para as análises subsequentes.

## 2.3 Reutilização de Funções

O código foi desenvolvido com foco na modularidade e reutilização. As funções implementadas atendem tanto ao fluxo de **backfill** quanto ao **incremental**, promovendo consistência no processamento e facilitando a manutenção do pipeline. Essa abordagem também permite futuras expansões com menor esforço de desenvolvimento, reduzindo a duplicidade de código e otimizando a gestão das rotinas de ingestão de dados.

### 3. Descrição das Transformações Aplicadas

O processo de transformação de dados foi projetado para garantir a consistência, integridade e qualidade das informações antes de serem salvos para verificação de hipóteses construção de dashboards.

#### 3.1 Tipagem e Conversões de Dados

Para assegurar a conformidade com as informações contidas nos dados foram realizados **casts** nos campos. Essa etapa garante que os dados estejam devidamente tipados e consistentes com as expectativas do banco de dados de destino. Dados iniciais:

- **id:** Inteiro (Integer) – Identificador único da transação.
- **won\_time:** Data (Datetime) – Data em que a venda foi realizada.
- **vendedor:** String (String) – Nome do vendedor responsável pela transação.
- **mrr\_receita\_recorrente:** Numérico (Numeric) – Valor referente à receita recorrente gerada pela venda.
- **nrr\_receita\_nao\_recorrente:** Numérico (Numeric) – Valor referente à receita não recorrente gerada pela venda.
- **pipeline\_id:** Inteiro (Integer) – Campo usado para rastrear o pipeline ou estágio no processo de vendas.
- **plano:** String (String) – Nome do plano associado à venda.
- **periodicidade:** String (String) – Frequência de pagamento ou renovação do plano (mensal, anual, etc.).
- **churn\_rate:** Numérico (Numeric) – Taxa de cancelamento (churn) associada à transação.

Após o processo de Feature Engineering novos dados foram criados e isso acarreta nas seguintes colunas com novas

- **ano:** Inteiro (Integer) – Ano extraído da data da venda.
- **mes:** Inteiro (Integer) – Mês extraído da data da venda.
- **dia:** Inteiro (Integer) – Dia extraído da data da venda.
- **ano\_mes:** Data (Datetime) – Representação consolidada do mês (AAAA-MM).
- **dia\_da\_semana:** Inteiro (Integer) – Código numérico do dia da semana (0 = segunda-feira).
- **semana\_do\_ano:** Inteiro (Integer) – Número da semana no ano.
- **nome\_mes:** String (String) – Nome do mês.
- **nome\_dia\_semana:** String (String) – Nome do dia da semana.
- **eh\_final\_de\_semana:** Booleano (Boolean) – Indicação se a venda ocorreu no fim de semana.
- **receita\_total:** Numérico (Numeric) – Soma do MRR e NRR.
- **proporcao\_mrr\_nrr:** Numérico (Numeric) – Proporção entre MRR e NRR.



- **mrr\_categoria:** String (String) – Classificação do MRR (baixo, médio, alto).
- **churn\_alto:** Booleano (Boolean) – Indicador se o churn está acima do limite definido.
- **qtd\_vendas\_totais:** Inteiro (Integer) – Total acumulado de vendas.
- **qtd\_vendas\_ano:** Inteiro (Integer) – Total de vendas no ano.
- **qtd\_vendas\_mensais:** Inteiro (Integer) – Total de vendas no mês.
- **qtd\_vendas\_semanal:** Inteiro (Integer) – Total de vendas na semana.
- **qtd\_tipos\_planos\_mes:** Inteiro (Integer) – Quantidade de tipos de planos vendidos no mês.
- **qtd\_tipos\_planos\_ano:** Inteiro (Integer) – Quantidade de tipos de planos vendidos no ano.
- **qtd\_tipos\_planos\_vendidos:** Inteiro (Integer) – Quantidade total de tipos de planos vendidos.
- **dias\_sem\_venda:** Numérico (Numeric) – Número de dias desde a última venda.
- **tempo\_medio\_entre\_vendas:** Numérico (Numeric) – Tempo médio entre vendas.
- **comissao\_percentual:** Numérico (Numeric) – Percentual aplicado no cálculo da comissão.
- **valor\_comissao:** Numérico (Numeric) – Valor da comissão gerada pela venda.
- **categoria\_desempenho:** String (String) – Classificação do desempenho do vendedor (ex.: abaixo\_da\_meta).

### 3.2 Limpeza de Dados

Com o objetivo de eliminar inconsistências e assegurar a integridade das informações, foi realizada uma etapa específica de remoção de linhas nulas e inválidas em campos essenciais para as análises e para a qualidade geral do dataset.

Os campos validados nessa etapa incluem:

- id
- won\_time
- vendedor
- mrr (receita recorrente)
- nrr(receita não recorrente)
- pipeline\_id
- plano
- periodicidade
- churn\_rate

Além da verificação de valores nulos, também foram identificadas e removidas linhas contendo entradas claramente inválidas, como registros preenchidos apenas com a palavra “deleted” ou linhas nas quais o valor de uma célula correspondia ao próprio nome da coluna (por exemplo, registros no campo won\_time contendo o texto “won\_time”). Essa etapa foi fundamental para garantir que apenas dados consistentes e confiáveis fossem utilizados nas análises subsequentes.

## 4. Análise Descritiva dos Dados

### 4.1 Dados Numericos

As variáveis numéricas presentes no conjunto de dados fornecem informações valiosas sobre os aspectos financeiros das transações e os comportamentos dos clientes, incluindo receitas recorrentes e não recorrentes, além da taxa de churn. Essas variáveis foram examinadas quanto ao tipo, diversidade de valores, distribuição e possíveis outliers, gerando insights que orientam tanto estratégias comerciais quanto análises de qualidade dos dados.

A seguir, detalhamos as análises realizadas para cada uma dessas variáveis.

- **mrr\_receita\_recorrente :**

A coluna `mrr_receita_recorrente` representa os valores recorrentes gerados pelas vendas e é composta por números inteiros não negativos. O conjunto apresentou 540 valores únicos, refletindo alta diversidade e baixa repetição, o que indica uma base rica para segmentação. A distribuição revela uma concentração marcante entre 200 e 400, com a frequência diminuindo à medida que os valores aumentam. Chamam atenção os registros no valor de 800, que aparecem repetidamente e podem representar limites máximos contratuais ou categorias premium. A densidade de dados confirma o pico entre 200 e 400, enquanto valores acima de 800 são extremamente raros. Como insights práticos, os clientes na faixa de MRR baixo (200–400) podem ser priorizados em estratégias de upsell, enquanto aqueles nos níveis altos (600–800) demandam atenção especial para retenção e fidelização. Outliers identificados, como os valores em 800, precisam ser avaliados quanto à sua legitimidade.

- **nrr\_receita\_nao\_recorrente :**

A coluna `nrr_receita_nao_recorrente` traz os valores únicos e pontuais gerados fora dos contratos recorrentes. Também composta por números inteiros não negativos, apresentou 389 valores únicos, indicando diversidade moderada, menor do que a observada no MRR. A distribuição mapeada por boxplots e gráficos de densidade mostrou que a maioria dos valores se concentra entre 0 e 300, com quedas suaves até 600. Os valores acima de 500 aparecem como potenciais outliers, recomendando investigação mais detalhada. Mesmo entre valores baixos, como 67 ou 114, observa-se heterogeneidade, possivelmente vinculada a diferentes tipos de serviços ou características específicas de clientes. Esses padrões permitem identificar oportunidades para aumentar o valor médio das transações pontuais e ajustar políticas comerciais.

- **churn\_rate :**

A coluna **churn\_rate** apresenta os índices de cancelamento associados às vendas, sendo composta por valores decimais. A análise revelou apenas 4 valores únicos, sugerindo baixa granularidade e possível categorização prévia em faixas específicas. A variação vai de 0,075 a 0,25, com maior concentração entre 0,10 e 0,15. Valores acima de 0,25, quando presentes, são críticos e podem sinalizar segmentos de alta rotatividade, demandando ações imediatas. A distribuição geral, apontada por histogramas e gráficos de densidade, reforça a pouca diversidade, o que limita a profundidade das análises, já que nuances finas entre, por exemplo, 0,11 e 0,14 acabam diluídas. Ainda assim, a informação é relevante para segmentação estratégica, identificação de riscos e priorização de esforços para reduzir o churn em grupos mais vulneráveis.

## 4.2 Dados Categóricos

As variáveis categóricas do conjunto de dados fornecem informações essenciais sobre o perfil das transações, como periodicidade dos contratos, responsáveis comerciais e os planos contratados. A análise dessas variáveis permite identificar padrões de comportamento, distribuição de cargas entre a equipe de vendas e a relevância estratégica de determinados produtos ou serviços. A seguir, detalhamos as análises realizadas para cada coluna categórica.

A seguir, detalhamos as análises realizadas para cada uma dessas variáveis.

- **periodicidade:**

A coluna **periodicidade** apresenta três categorias distintas: Mensal, Anual e Semestral. Com um total de 3 valores únicos, ela revela forte predominância do modelo mensal, que representa aproximadamente 84,5% das transações (845 ocorrências em um total estimado de 999 registros). O modelo anual aparece em segundo lugar, com 14,4% (144 ocorrências), enquanto o semestral responde por apenas 1% (10 ocorrências). Essa distribuição segue nitidamente o princípio de Pareto, onde Mensal e Anual somam cerca de 98,9% dos casos, deixando o Semestral com contribuição marginal. O domínio absoluto do plano mensal indica uma preferência clara do mercado por contratos de curto prazo, enquanto a baixa presença do modelo semestral sugere que ele pode atender apenas nichos específicos ou situações sazonais. Recomenda-se investigar a legitimidade desses registros menos representativos e avaliar oportunidades de expansão para outros modelos de periodicidade.

- **vendedor:**

A coluna **vendedor** contém os nomes dos profissionais responsáveis pelas vendas, totalizando 5 valores únicos. A distribuição é fortemente concentrada em Igor Martins, que responde por cerca de 63,9% das transações (639 ocorrências), seguido por Rafael Cunha (12,6%), Fernanda Dias (11,5%), Bruna Lopes (10,3%) e Larissa Melo (1,6%). Aqui, o princípio de Pareto também se aplica: os três primeiros vendedores acumulam aproximadamente 88% das vendas, destacando-os como peças-chave para o desempenho comercial da empresa. O claro desequilíbrio, com um único vendedor liderando mais de 60% das vendas, levanta questões sobre dependência excessiva e possíveis riscos operacionais. Já a baixa presença de Larissa Melo pode indicar que ela seja nova na equipe, esteja focada em segmentos específicos ou enfrente desafios de desempenho. A equipe enxuta, composta por apenas cinco nomes, reforça a importância de monitorar cuidadosamente a distribuição de cargas e identificar oportunidades para equilibrar resultados entre os membros.

- **plano:**

A coluna **plano** contém os nomes dos profissionais responsáveis pelas vendas, totalizando 5 valores únicos. A distribuição é fortemente concentrada em Igor Martins, que responde por cerca de 63,9% das transações (639 ocorrências), seguido por Rafael Cunha (12,6%), Fernanda Dias (11,5%), Bruna Lopes (10,3%) e Larissa Melo (1,6%). Aqui, o princípio de Pareto também se aplica: os três primeiros vendedores acumulam aproximadamente 88% das vendas, destacando-os como peças-chave para o desempenho comercial da empresa. O claro desequilíbrio, com um único vendedor liderando mais de 60% das vendas, levanta questões sobre dependência excessiva e possíveis riscos operacionais. Já a baixa presença de Larissa Melo pode indicar que ela seja nova na equipe, esteja focada em segmentos específicos ou enfrente desafios de desempenho. A equipe enxuta, composta por apenas cinco nomes, reforça a importância de monitorar cuidadosamente a distribuição de cargas e identificar oportunidades para equilibrar resultados entre os membros.

## 4.3 Dados Temporais

As variáveis temporais, em especial a coluna `won_time`, oferecem insights fundamentais sobre a dinâmica temporal das vendas, revelando padrões sazonais, quedas inesperadas e possíveis falhas operacionais. Uma análise cuidadosa dessas informações ajuda a identificar períodos de pico, lacunas de registro e tendências ao longo do tempo. A seguir, detalhamos as observações extraídas das análises aplicadas a essa variável.

Ao observar a frequência mensal de `won_time`, nota-se que os meses de janeiro a agosto de 2022 concentraram os maiores volumes de vendas, com picos claros em janeiro, abril e maio. A partir de setembro de 2022, o volume mensal colapsa, mantendo-se extremamente baixo até agosto de 2023. Esse declínio acentuado reforça a necessidade de validação: trata-se de uma mudança estrutural no negócio? Houve alteração no pipeline de ingestão de dados? Ou os registros simplesmente não foram carregados no período posterior? Essa informação é essencial para garantir análises temporais precisas.

A análise da **distribuição diária de `won_time`** usando boxplots mostra como os registros estão distribuídos ao longo dos dias do mês (1 a 31). A dispersão é relativamente equilibrada, com registros espalhados por praticamente todos os dias, mas observa-se uma leve concentração nos dias iniciais e médios do mês (1–15). Além disso, surgem alguns outliers, representando dias atípicos com volumes muito elevados (acima de 10–12 vendas). A mediana diária varia de 3 a 6 registros, indicando uma estabilidade moderada, mas não completamente uniforme ao longo do ciclo mensal. Esses padrões ajudam a identificar se há sazonalidade operacional interna ou se picos específicos merecem investigação detalhada.

## 5. Hipóteses e Padrões Identificados

A validação de hipóteses foi conduzida para entender melhor as relações entre variáveis-chave do conjunto de dados e identificar insights acionáveis para estratégias comerciais. Nesta seção, detalhamos a análise da primeira hipótese investigada.

### 5.1 Hipótese 1: Relação entre Desempenho do Vendedor e Receita Total

A hipótese parte do princípio de que o desempenho individual dos vendedores, medido pelo atingimento de metas, impacta diretamente a quantidade de vendas realizadas, a receita média gerada por venda e a receita total acumulada.

A análise dos dados revelou que, em termos de quantidade de vendas por categoria de desempenho, os vendedores classificados como abaixo da meta apresentam o menor número de transações registradas, enquanto aqueles próximos da meta ocupam uma posição intermediária. Os vendedores que atingiram a meta lideram em volume de vendas, destacando-se como os principais contribuintes para os resultados comerciais.

Na análise da receita média por categoria, observou-se uma progressão clara: vendedores abaixo da meta apresentam uma receita média estimada em R\$ 200, vendedores próximos da meta alcançam cerca de R\$ 400, e aqueles que atingiram a meta chegam a R\$ 700. Esse padrão se reflete também na receita total por categoria, com registros abaixo da meta concentrando-se em faixas de até R\$ 200, próximos da meta movimentando entre R\$ 400 e R\$ 800, e os que atingiram a meta superando frequentemente R\$ 1.000, com valores mais altos e dispersos.

As principais conclusões indicam uma correlação positiva robusta: vendedores que atingem a meta não apenas geram mais vendas, mas também produzem receitas médias e totais significativamente superiores. Um exemplo claro é que a categoria atingiu a meta apresenta uma receita total aproximadamente 3,5 vezes maior que a categoria abaixo da meta.

Do ponto de vista estratégico, a análise recomenda atenção especial aos vendedores classificados como próximos da meta, pois eles estão em uma zona de alavancagem: intervenções direcionadas, como treinamento adicional ou programas de incentivo, podem rapidamente transformá-los em alto desempenho. Já a categoria abaixo da meta requer uma investigação detalhada das causas, que podem estar ligadas a fatores como capacitação insuficiente, baixa motivação ou desafios operacionais específicos.

## 5.2 Hipótese 2: Planos com Periodicidade Anual Têm Menor Churn Rate?

A segunda hipótese investigada busca entender se há uma relação entre a periodicidade do plano contratado e a taxa de cancelamento (churn rate), assumindo que planos anuais, por envolverem compromissos mais longos, oferecem maior estabilidade e retenção de clientes.

A análise inicial revelou que o churn rate médio por periodicidade confirma essa expectativa: planos anuais apresentam a menor taxa, com 0,08 (8%), seguidos pelos semestrais com 0,10 (10%) e, por fim, pelos mensais com 0,15 (15%). Isso já sugere que, proporcionalmente, clientes de contratos anuais permanecem ativos por mais tempo.

Quando examinada a distribuição do churn rate, observa-se que os planos anuais concentram seus valores abaixo de 0,10, com baixa dispersão, reforçando um padrão de estabilidade. Por outro lado, planos mensais apresentam churn rates espalhados entre 0,10 e 0,25, refletindo maior rotatividade. Os planos semestrais se posicionam de forma intermediária, com churn rates entre 0,10 e 0,15, mas a análise aqui requer cautela devido ao tamanho reduzido da amostra.

Em termos de quantidade de vendas por periodicidade, a distribuição mostra forte predominância de planos mensais (845 vendas, ou 84,5% do total), enquanto planos anuais somam apenas 144 vendas (14,4%) e os semestrais apenas 10 vendas (1%). Isso significa que, embora os planos anuais apresentem melhor retenção, representam uma fatia relativamente pequena da base de clientes, o que limita seu impacto geral nos resultados globais.

As principais conclusões indicam claramente que os planos anuais possuem o churn rate mais baixo, possivelmente por oferecerem vantagens como descontos, benefícios exclusivos ou simplesmente pelo compromisso de longo prazo firmado no contrato. Já os planos mensais, por sua natureza flexível e de curto prazo, tendem a registrar maior rotatividade. No caso dos planos semestrais, o número extremamente baixo de vendas levanta um alerta: pode haver baixa atratividade comercial ou até desconhecimento dessa opção por parte dos clientes, tornando difícil tirar conclusões sólidas apenas com os dados disponíveis.

### **5.3 Hipótese 3: Vendedores com Maior Diversificação de Planos Vendidos Têm Maior Receita Total e Menor Churn Rate?**

A terceira hipótese explora se vendedores que atuam com uma gama mais ampla de planos vendidos — medida pela variável `categoria_diversificacao` — conseguem alcançar resultados superiores em termos de receita total e retenção de clientes (menor churn rate). A expectativa inicial era de que maior diversificação permitiria atender diferentes perfis de clientes, gerar mais receita e, consequentemente, reduzir cancelamentos.

A análise da distribuição de `qtd_tipos_planos_vendidos` mostrou que a maioria absoluta dos vendedores já opera com 6 tipos de planos, enquanto apenas alguns trabalham com 4. Isso significa que, ao contrário do que se poderia imaginar, a alta diversificação não é um fator raro ou diferencial — ela já é o padrão predominante na equipe.

Ao observar os boxplots de receita total por nível de diversificação, percebe-se que os vendedores com alta diversificação apresentam uma mediana ligeiramente superior em receita total quando comparados aos de baixa diversificação. No entanto, as distribuições se sobrepõem consideravelmente, com grande amplitude interquartil e presença de outliers em ambos os grupos. Isso indica que nem sempre ter um portfólio diversificado garante resultados excepcionais, já que há alta variabilidade de desempenho dentro de cada categoria, possivelmente influenciada por outros fatores não diretamente ligados à diversificação.

Surpreendentemente, a análise do churn rate médio por diversificação de planos revelou que os vendedores classificados com Alta Diversificação apresentam um churn médio mais alto (~10%) do que os de Baixa Diversificação (~8%). Esse resultado contraria a hipótese inicial, sugerindo que oferecer muitos tipos de planos não necessariamente fideliza clientes — e pode até gerar complexidade operacional que aumenta a rotatividade.

Os resultados foram confirmados quantitativamente pela matriz de correlação, onde a relação entre `nivel_diversificacao` e `receita_total` foi praticamente nula (0,0041) e a correlação entre `nivel_diversificacao` e `churn_rate` foi fraca e positiva (0,11). Isso reforça a conclusão de que, no conjunto analisado, a diversificação de portfólio não apresenta relação estatística relevante nem com receita total nem com churn rate.

Com base nesses achados, a hipótese foi rejeitada: não há evidências de que vendedores com maior diversificação de planos vendidos obtenham desempenho financeiro superior ou retenham clientes de forma mais eficaz.

Insights importantes surgem dessa análise: estratégias comerciais não devem se apoiar apenas na diversificação do portfólio. É necessário priorizar a qualidade das vendas, garantir a adequação dos planos ao perfil de cada cliente e implementar práticas ativas de gestão de churn, visando não apenas expandir o leque de produtos ofertados, mas também maximizar o valor gerado por cada venda.



## 6. Observações Gerais e Insights

A análise detalhada do conjunto de dados revelou padrões importantes tanto nas dimensões numéricas quanto categóricas, além de validar (ou rejeitar) hipóteses-chave que ajudam a entender o desempenho comercial, a estrutura do portfólio e os desafios relacionados à retenção de clientes.

Primeiramente, foi observado que a base de dados apresenta boa diversidade em variáveis como receitas (`mrr_receita_recorrente` e `nrr_receita_nao_recorrente`), embora concentre a maior parte dos registros em faixas médias e baixas, sugerindo que há espaço para explorar estratégias de upsell. Em contrapartida, a variável `churn_rate` demonstrou baixa granularidade, com apenas quatro valores distintos, o que limita análises mais sofisticadas e pode indicar um processo de categorização prévia que merece ser revisitado para aumentar o detalhamento das análises futuras.

Nas variáveis categóricas, percebeu-se forte predominância de planos mensais e de vendedores específicos, com Igor Martins liderando isoladamente em vendas, e o plano Astrea Starter concentrando quase 70% dos contratos. Esses desequilíbrios sugerem dependência excessiva de poucos produtos e pessoas-chave, o que representa risco estratégico caso haja mudanças inesperadas (ex.: saída de talentos ou alterações no mercado). Além disso, verificou-se que, apesar do baixo volume de vendas de planos anuais, esses contratos têm `churn_rate` significativamente menor, reforçando que estratégias de retenção podem ser fortalecidas ao incentivar compromissos de longo prazo.

Quanto às hipóteses analisadas, um dos insights mais relevantes foi que a diversificação de portfólio entre os vendedores não está associada a maior receita nem a menor rotatividade. Isso desafia o senso comum de que “vender mais tipos de produto” automaticamente gera melhores resultados e sugere que a qualidade e o foco das vendas importam mais do que o simples volume de ofertas.

Por fim, foi identificada uma possível lacuna temporal nos dados após setembro de 2022, o que pode refletir falhas na ingestão, encerramento de operações ou apenas um recorte temporal aplicado na extração — ponto que deve ser esclarecido para evitar conclusões enviesadas.

## 6.1 Principais Recomendações

- Reforçar estratégias de upsell para clientes nos planos de entrada (200–400 de MRR), aproveitando o espaço de crescimento.
- Incentivar a migração para planos anuais, aproveitando seu menor churn rate.
- Investigar as causas do alto churn em vendedores com alta diversificação e rever práticas comerciais associadas.
- Revisar a granularidade das métricas categorizadas (como `churn_rate`), permitindo análises mais detalhadas no futuro.
- Validar e corrigir (se necessário) possíveis lacunas ou rupturas temporais nos dados para assegurar a integridade das análises históricas.

## 7. Conclusão

A análise conduzida neste trabalho forneceu uma visão ampla e detalhada sobre os dados comerciais, destacando padrões relevantes, confirmando algumas suposições estratégicas e, ao mesmo tempo, desafiando outras percepções prévias. Observamos que os resultados financeiros e de retenção não dependem exclusivamente de volume ou diversificação, mas estão fortemente ligados à qualidade das operações, ao foco em segmentos estratégicos e à adequação dos produtos ao perfil dos clientes.

A predominância de planos mensais e de determinados vendedores, embora atualmente sustente a maior parte do desempenho, representa tanto uma fortaleza quanto uma vulnerabilidade: esses pilares precisam ser fortalecidos com atenção contínua, mas também devem ser equilibrados com iniciativas que ampliem a base de produtos e talentos para reduzir riscos de concentração.

Além disso, ficou claro que há oportunidades claras para otimização, como incentivar a adesão a planos de longo prazo, revisar métricas agregadas que limitam as análises (como o churn agrupado) e investigar possíveis falhas ou cortes temporais nos registros históricos.

Este estudo não apenas oferece um diagnóstico detalhado da situação atual, mas também aponta caminhos estratégicos para melhorar a performance futura, aumentar a resiliência das operações e aprofundar a capacidade analítica da organização.