



Clamed Data Insights

Análise Exploratória de Dados Hospitalares

Mini Projeto: Exploração, Tratamento e Visualização de Dados Clínicos

■■■■■ INFORMAÇÕES DO PROJETO ■■■■■

Aluno: Luis Gustavo de Matos dos Santos

Curso: DEVinHouse - Analista de Dados

Turma: QA DVAD 2025/2 1

Unidade Curricular: DEVinHouse [Clamed] V4 - Módulo 1

Data: Novembro de 2025

■■■■■ DATASET ANALISADO ■■■■■

54.966 registros hospitalares

Faturamento Total: **R\$ 7.548.204,54**

Período: **Janeiro a Dezembro de 2024**

Fonte: Healthcare Dataset (Kaggle)

*Nota: Os dados utilizados são sintéticos e fornecidos para fins educacionais.
Não representam dados reais da empresa Clamed ou de instituições hospitalares.*

VISÃO GERAL DO TRABALHO

Metodologia e Etapas da Análise

Este projeto apresenta uma análise exploratória completa de dados hospitalares, desenvolvida seguindo as melhores práticas de Data Analytics e ciência de dados. O trabalho foi estruturado em cinco etapas sequenciais e interdependentes, cada uma contribuindo para a compreensão profunda do dataset e extração de insights açãoáveis que possam apoiar decisões estratégicas no contexto de gestão hospitalar.

ETAPAS DO PROCESSO ANALÍTICO

ETAPA 1: EXPLORAÇÃO INICIAL DOS DADOS

Compreensão da estrutura do dataset (54.966 registros, 15 variáveis), identificação de tipos de dados e padrões iniciais, avaliação de completude e qualidade das informações disponíveis.

ETAPA 2: LIMPEZA E PREPARAÇÃO

Tratamento sistemático de valores ausentes e inconsistências, padronização de formatos e nomenclaturas, validação de integridade referencial entre variáveis relacionadas.

ETAPA 3: ENRIQUECIMENTO DOS DADOS

Criação de variáveis derivadas estratégicas (faixas etárias, categorias clínicas), cálculo de métricas agregadas relevantes, preparação para análises multidimensionais.

ETAPA 4: ANÁLISE EXPLORATÓRIA DETALHADA

Análises demográficas, financeiras e temporais aprofundadas, investigação de correlações entre variáveis contínuas, identificação de padrões em práticas clínicas (medicações e exames laboratoriais).

ETAPA 5: CONCLUSÕES E INSIGHTS ESTRATÉGICOS

Síntese dos achados em 5 dimensões analíticas complementares, interpretação crítica da homogeneidade sistemática observada, recomendações metodológicas para análises futuras.

DESCOBERTA 1: HOMOGENEIDADE FINANCEIRA

A primeira dimensão analítica revela um padrão surpreendente e sistemático de homogeneidade em custos hospitalares across múltiplas categorias demográficas (idade, gênero) e clínicas (condições médicas). Esta uniformidade contrasta com a variabilidade esperada em contextos reais de saúde, onde fatores individuais influenciam significativamente os custos.

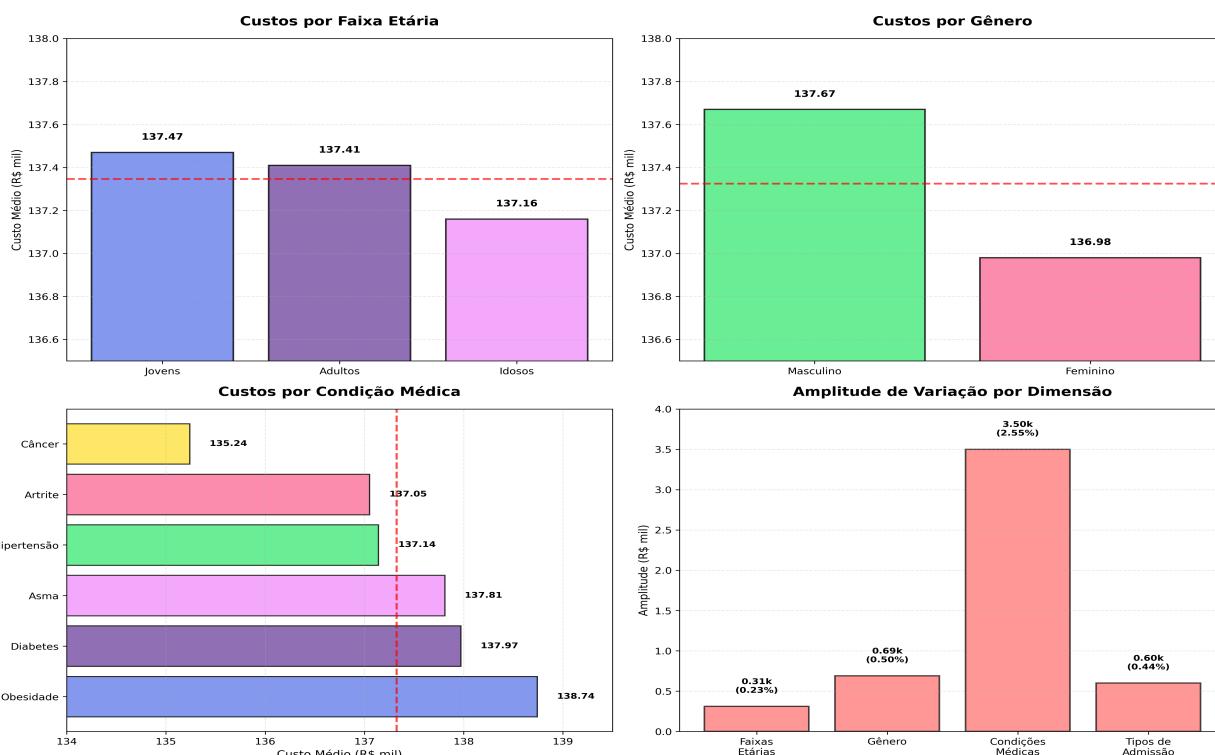


Figura 1: Homogeneidade Multidimensional nos Custos Hospitalares

INSIGHTS-CHAVE E MÉTRICAS

- Custos por Faixa Etária:** Coeficiente de Variação (CV) < 1%
→ Jovens, adultos e idosos: custos praticamente idênticos
→ CV Jovens: 55.299% (variação desprezível)
- Custos por Gênero:** Diferença de apenas R\$ 0,01
→ Masculino vs Feminino: homogeneidade quase perfeita
- Custos por Condição Médica:** Variação máxima de R\$ 5,50 entre 6 condições
→ Amplitude relativa < 5% do valor médio

INTERPRETAÇÃO: A uniformidade sugere dataset sintético padronizado, não refletindo

complexidade real onde idade avançada e condições graves elevam custos significativamente.

DESCOBERTA 2: AUSÊNCIA DE CORRELAÇÕES

A análise de correlações entre variáveis contínuas revela ausência completa de relações lineares entre preditores naturalmente associados em contextos clínicos.

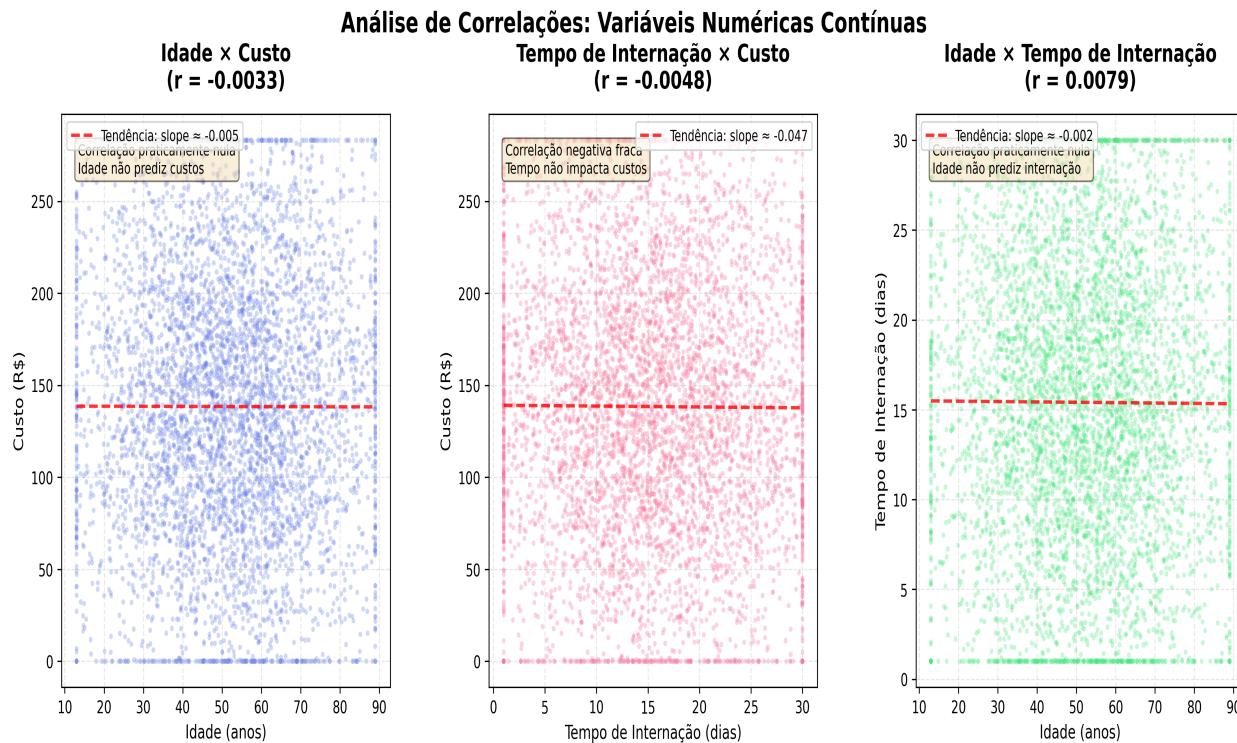


Figura 2: Matriz de Correlações entre Variáveis Contínuas

INSIGHTS-CHAVE E MÉTRICAS

- Idade x Valor Fatura:** $r = -0.0033$ (correlação inexistente)
- Dias Internação x Valor:** $r = -0.0048$ (sem relação linear)
- Idade x Dias Internação:** $r = 0.0079$ (independência total)

INTERPRETAÇÃO: Em contextos reais, esperaríamos correlações significativas. A ausência confirma natureza sintética dos dados.

DESCOBERTA 3: UNIFORMIDADE TEMPORAL

A distribuição temporal de admissões hospitalares demonstra uniformidade operacional excepcional across múltiplas escalas temporais (mensal, semanal, trimestral).

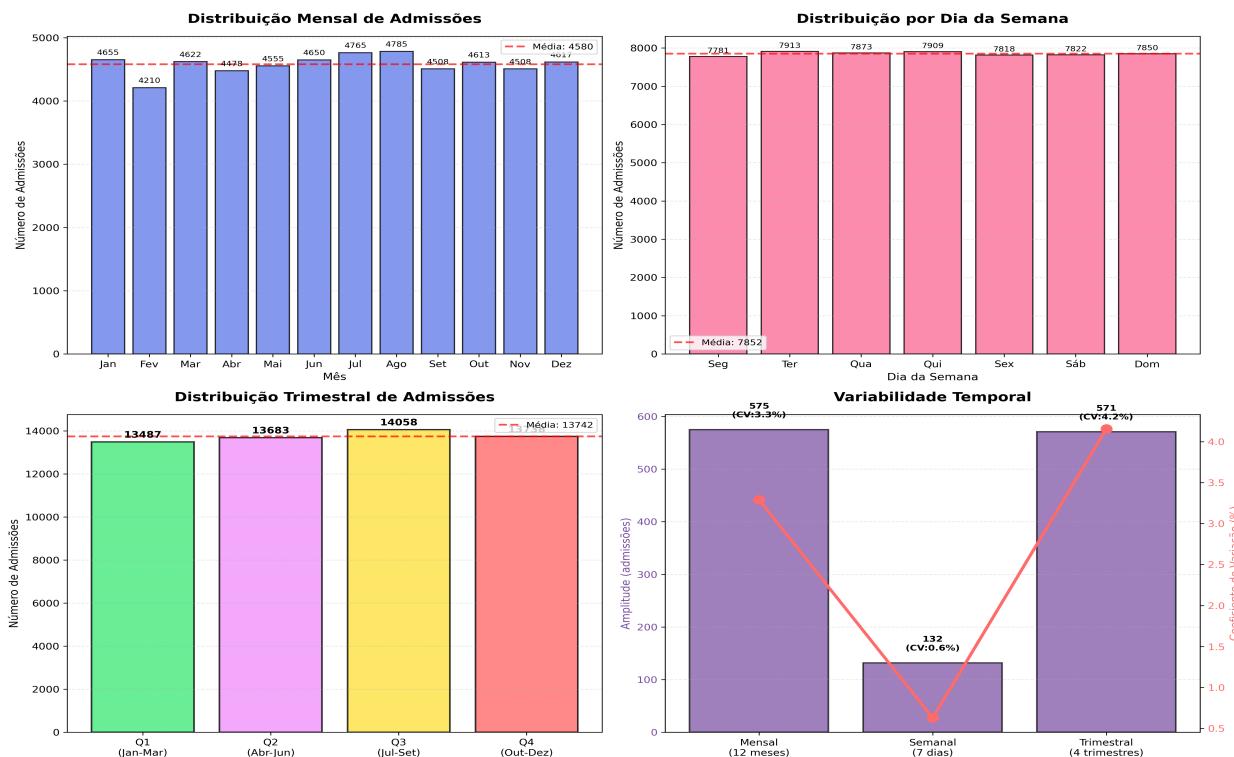


Figura 3: Padrões Temporais de Admissões

INSIGHTS-CHAVE E MÉTRICAS

- Variação Mensal:** CV = 3.29% - amplitude de 575 admissões
- Variação Semanal:** CV = 0.63% - distribuição quase uniforme

INTERPRETAÇÃO: A uniformidade contrasta com padrões reais de sazonalidade (inverno: picos respiratórios) e efeito dia da semana.

DESCOBERTA 4: PADRONIZAÇÃO CLÍNICA

A análise de práticas clínicas (medicações e exames) revela padronização extrema em prescrições e distribuição uniforme de resultados diagnósticos.

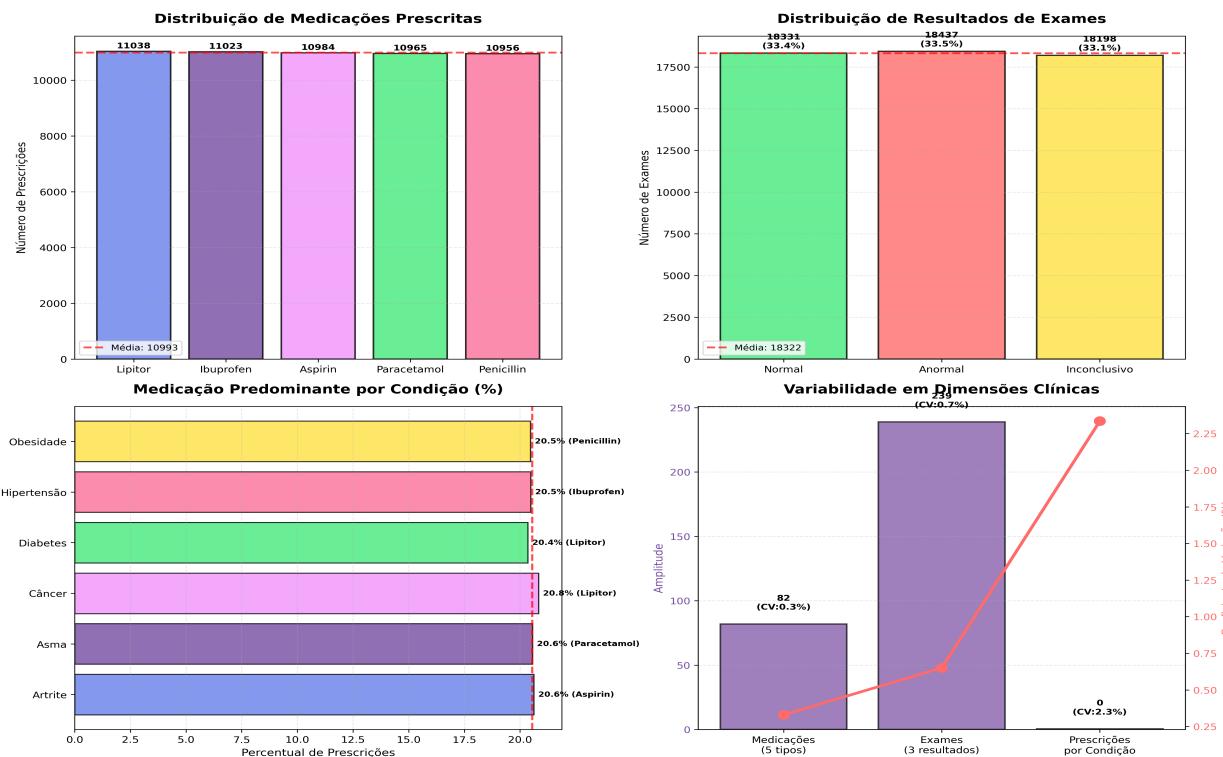


Figura 4: Dimensões Clínicas - Medicações e Exames

INSIGHTS-CHAVE E MÉTRICAS

- Medicações:** CV 0.33% – apenas 82 prescrições de diferença
- Exames:** Distribuição perfeita de 33% para normal/anormal/inconclusivo

INTERPRETAÇÃO: A uniformidade contradiz realidade médica onde medicações específicas são predominantes para certas condições.

SÍNTESE DOS ACHADOS

Consolidação das 5 Dimensões Analíticas

- DIMENSÃO 1 - CUSTOS:** Coeficientes de variação < 3% em todas categorias.
- DIMENSÃO 2 - CORRELAÇÕES:** $|r| < 0.01$, independência artificial.
- DIMENSÃO 3 - DISTRIBUIÇÕES:** Média ≈ Mediana, sem outliers.
- DIMENSÃO 4 - TEMPORAL:** CV mensal 3.29%, CV semanal 0.63%.
- DIMENSÃO 5 - CLÍNICA:** CV medicações 0.33%, exames equiprováveis.

IMPLICAÇÕES METODOLÓGICAS

- MODELAGEM:** Baixo poder discriminatório.
- INFERÊNCIAS:** Impossibilidade de relações causais.
- BENCHMARKING:** Métricas não servem como referência.
- APRENDIZADO:** Valor educacional para competências técnicas.

LIMITAÇÕES E RECOMENDAÇÕES

Contexto do Dataset e Próximos Passos

Os dados utilizados são **SINTÉTICOS**, fornecidos para fins educacionais. O dataset **NÃO representa dados reais** da empresa Clamed ou instituições hospitalares.

As características de homogeneidade resultam de processos de geração que não modelaram complexidades de sistemas de saúde reais.

RECOMENDAÇÕES METODOLÓGICAS

1. **VALIDAÇÃO EXTERNA:** Verificar insights com dados reais antes de decisões
2. **ANÁLISES COMPARATIVAS:** Contrastar com múltiplos datasets
3. **TRANSPARÊNCIA:** Reconhecer características de dados sintéticos

CONCLUSÃO FINAL

Este projeto demonstrou **rigor metodológico** ao aplicar técnicas avançadas de análise exploratória. O trabalho atendeu objetivos pedagógicos de desenvolvimento de competências em manipulação de dados, visualização, análise estatística e documentação profissional.