

# Relatório de Análise Exploratória dos dados (EDA)

Projeto: Painel Gerencial de Auditória- A.M.I Saúde

Responsável: Gustavo A. André

Portifólio completo: [GitHub](#)

Bases externas utilizadas: [CID-10 Capítulo](#)

## 1. Avaliação da integridade das Bases:

A primeira etapa constituiu na auditoria de qualidade das duas bases fornecidas para garantir a confiabilidade dos indicadores estratégicos.

- Cardinalidade e cruzamentos: As bases apresentam uma relação 1:N via `senha_internacao`, onde uma internação possui múltiplos itens faturados.
- Saneamento de chaves: Foi identificado 10 registros de chaves com formato padrão (`SI[ANO][SEQUENCIAL]`) fora do esperado.
- Dados Faltantes: Identificamos que 6% dos registros das internações(1200) não possuem motivos de alta e nem data de alta, indicando pacientes ainda hospitalizados. A presença de 55% de valores nulos no CID secundário limita análises clínicas mais refinadas e pode reduzir a performance de modelos preditivos baseados em comorbidades.

## 2. Inconsistências Detectadas e regras de validação:

Para assegurar a governança dos dados, foram aplicadas as seguintes regras de validação:

- Validação temporal:
  - Verificação de cronologia onde a data de admissão é posterior data de alta. Foram encontrados 15 registros.
  - Foi detectado 20 registros com o tempo de autorização divergente do cálculo de tempo de autorização; `['data_autorizacao_senha'] - ['data_solicitacao_autorizacao']`
- Integridade Financeira:
  - Foi validado se há algum registro de valores negativos e não foi detectado nas bases.
  - Após auditoria dos valores de quantidade de itens \* valor unitário de itens, foi verificado uma inconsistência de centavos, necessita revisão de arredondamento de valores.
  - Foi identificado em 817 registros valores glosados superiores ao valor total do item, caracterizando uma anomalia financeira relevante que pode indicar falha de parametrização, erro operacional ou inconsistência de regra de glosa.
  - 1248 registros com o motivo glosa = Quantidade acima do padrão está igual ao valor total de itens, o que poderia ser menor, pois se trata de quantidades maiores de itens.
  - 1042 registros de itens duplicados estão menores que o valor total do item.
- Tipagem:
  - Todas as bases foram tratadas para que as colunas estivessem com as tipagens corretas.
- Validação dos CIDS:
  - Todos os Cids estão no formato esperado.
  - Todos os Cids pertencem a uma categoria existente na base extensa CID-10.

- Para futuras auditorias foi criado uma FLAG cid\_principal\_compativel\_especialidade para verificar se o CID pertence a especialidade que o paciente está tratando.
- Foi criado um alerta\_swap para verificar os seguintes pontos nos CIDs:
  - Cid secundário mais compatível com a especialidade do que o Cid primário
  - Cids genéricos não deveriam ser Cids principais, se houver um cid secundário mais específico, isso é um padrão de hierarquia invertida.
  - Cids Obstétricos e Piscíquiatricos deveriam ser o cid principal não o secundário, encontrados 64 registros como cids secundários.
  - Os scores criados são:
    - 0- Sem indicio de swap de cid
    - 1- Suspeita de swap de cid
    - 2- Alta probabilidade de inversão de cid

### 3. Principais achados e relações estatísticas:

A exploração revelou padrões críticos para a tomada de decisão da gerência executiva.

- Perfil de Permanência (LOS): A média de permanência hospitalar está fortemente correlacionada ao nível de complexidade, servindo como base para o acompanhamento da permanência hospitalar exigido no desafio.
- Driver de Custo (UTI): Através de um teste estatístico (T de Student), comprovou-se que internações com uso de UTI possuem custo médio 65% superior (R\$ 76.419 vs R\$ 46.318), com significância estatística ( $p < 0.05$ ).
- Eficiência Regulatória: O tempo\_autorizacao\_horas médio foi identificado como um KPI de SLA para a operação, especialmente em casos de caráter "Urgência".

De forma geral, os achados indicam que a operação apresenta boa consistência estrutural, porém com pontos críticos concentrados em regras de glosa, classificação diagnóstica e eficiência regulatória. Esses fatores impactam diretamente custo, previsibilidade financeira e qualidade assistencial, justificando a construção de um painel gerencial orientado à priorização de internações e monitoramento contínuo da operação.

### 4. Motivação para Modelagem Analítica:

A análise exploratória evidenciou padrões relevantes associados a risco operacional, como permanência hospitalar prolongada, concentração de custos em internações de maior complexidade, recorrência de glosas financeiras e ineficiências no processo regulatório. Adicionalmente, observou-se a ausência de um rótulo explícito que caracterize internações críticas, o que motivou a construção de variáveis proxy baseadas em critérios assistenciais, financeiros e regulatórios. Esses achados fundamentaram o desenvolvimento posterior de um score analítico de priorização de internações, com o objetivo de apoiar a tomada de decisão gerencial e o direcionamento de esforços de auditoria e acompanhamento assistencial.