

## **Centro Universitário Nobre – UNIFAN**

Alunos: Athos, Erika, Giulia

Bacharelado: Engenharia de software

Data: 29/05/2025

## **Relatório Técnico**

### **Projeto: Análise de Agrupamento e Reajustes em Planos de Saúde**

O projeto tem como objetivo analisar dados históricos de percentuais de reajuste e agrupamento de operadoras de planos de saúde, com base na base pública da ANS (Agência Nacional de Saúde Suplementar). A proposta visa identificar padrões entre operadoras e beneficiários e compreender como os reajustes se distribuem entre diferentes grupos, utilizando técnicas de Inteligência Artificial e visualização de dados.

### **Tecnologias Utilizadas**

- Linguagem: Python
- Manipulação e Análise de Dados: Pandas, Numpy
- Visualização de Dados: Matplotlib, Seaborn
- Modelagem Preditiva e Agrupamento: Scikit-learn (K-means, Regressão Linear)
- Fonte dos Dados: ANS - Dados Abertos  
(<https://dados.gov.br/dados/conjuntos-dados/percentuais-de-reajuste-de-agrupamento>)

### **Funcionamento do Sistema**

#### **1. Coleta de Dados:**

Os dados foram obtidos da base aberta da ANS e contém informações sobre operadoras, reajustes, número de beneficiários, entre outros.

#### **2. Limpeza e Preparação:**

- Remoção de dados duplicados

- Tratamento de valores ausentes
- Conversão de tipos de dados para formatos apropriados

### 3. Análises Realizadas:

- Método do Cotovelo: Determinação do número ideal de clusters (3)
- Clusterização K-means: Agrupamento das operadoras com base nos percentuais de reajuste
- Distribuição de Beneficiários: Análise da concentração por cluster
- Histogramas: Frequência de variáveis relevantes
- Matriz de Correlação: Relações entre percentuais e número de beneficiários
- Regressão Linear:
  - Real vs. Predito
  - Beneficiários vs. Percentual Único

### 4. Principais Resultados (com suporte gráfico):

- A maioria das operadoras (~5.000) está concentrada no Cluster 0
- Cluster 0 possui maiores percentuais ambulatoriais
- Os percentuais únicos não apresentam forte correlação com o número de beneficiários
- Cluster 1 e 2 representam comportamentos atípicos

## Objetivos Atendidos

- Identificar agrupamentos relevantes entre operadoras
- Visualizar padrões de reajuste e distribuição de beneficiários
- Avaliar a eficácia de técnicas de clusterização no setor da saúde
- Apoiar a tomada de decisão com base em dados públicos

## Como Executar o Projeto

### 1. Instale as dependências:

```
pip install -r requirements.txt
```

### 2. Execute o script principal:

```
python analise_reajustes.py
```

3. Os gráficos e resultados serão gerados automaticamente em formato de imagem e/ou exibidos em tela.

## **Conclusão**

A análise evidencia a existência de padrões relevantes entre operadoras de planos de saúde, destacando o potencial de ferramentas de clusterização e regressão para o entendimento do mercado suplementar. A predominância do Cluster 0 aponta para a uniformidade de reajustes baixos entre a maioria das operadoras, com implicações diretas sobre os beneficiários.

## **Importância da Informação**

O conhecimento gerado pode embasar ações da ANS e de políticas públicas que visem à regulação mais justa e eficiente dos reajustes nos planos de saúde. Entender a concentração de beneficiários em clusters com baixos reajustes ajuda a identificar grupos vulneráveis, promovendo transparência, equidade e planejamento estratégico para o setor.

## **1. Objetivo**

Analisar os dados históricos de percentuais de reajuste e agrupamento de operadoras de planos de saúde, com base na pesquisa anual da ANS.

## **2. Introdução**

Este relatório examina a distribuição de operadoras e beneficiários, bem como os resultados do método do cotovelo e regressão linear, com dados atualizados até 15 de maio de 2025.

## **3. Metodologia**

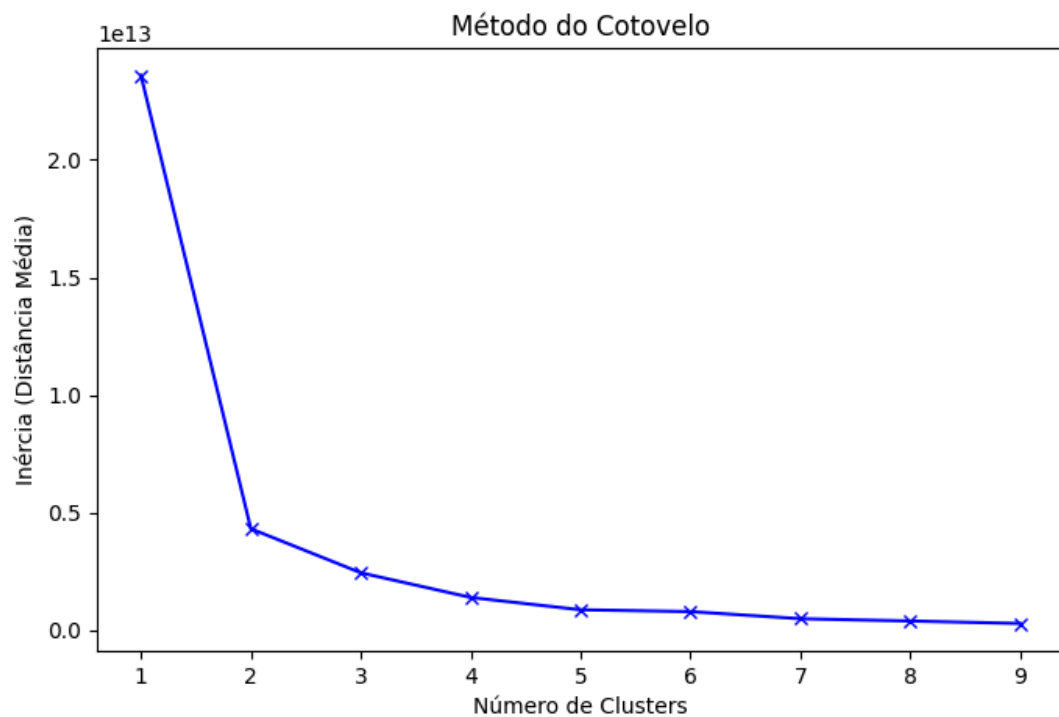
Os dados brutos foram obtidos do portal de dados abertos ao público, disponíveis em <https://dados.gov.br/dados/conjuntos-dados/percentuais-de-reajuste-de-agrupamento>, mantido pela ANS. Esses dados foram limpos utilizando um script em Python para

remover inconsistências, valores ausentes e duplicatas, garantindo a qualidade da análise. Posteriormente, os dados foram analisados por meio do método do cotovelo, clusterização K-means, histogramas e regressão linear.

## 4. Resultados

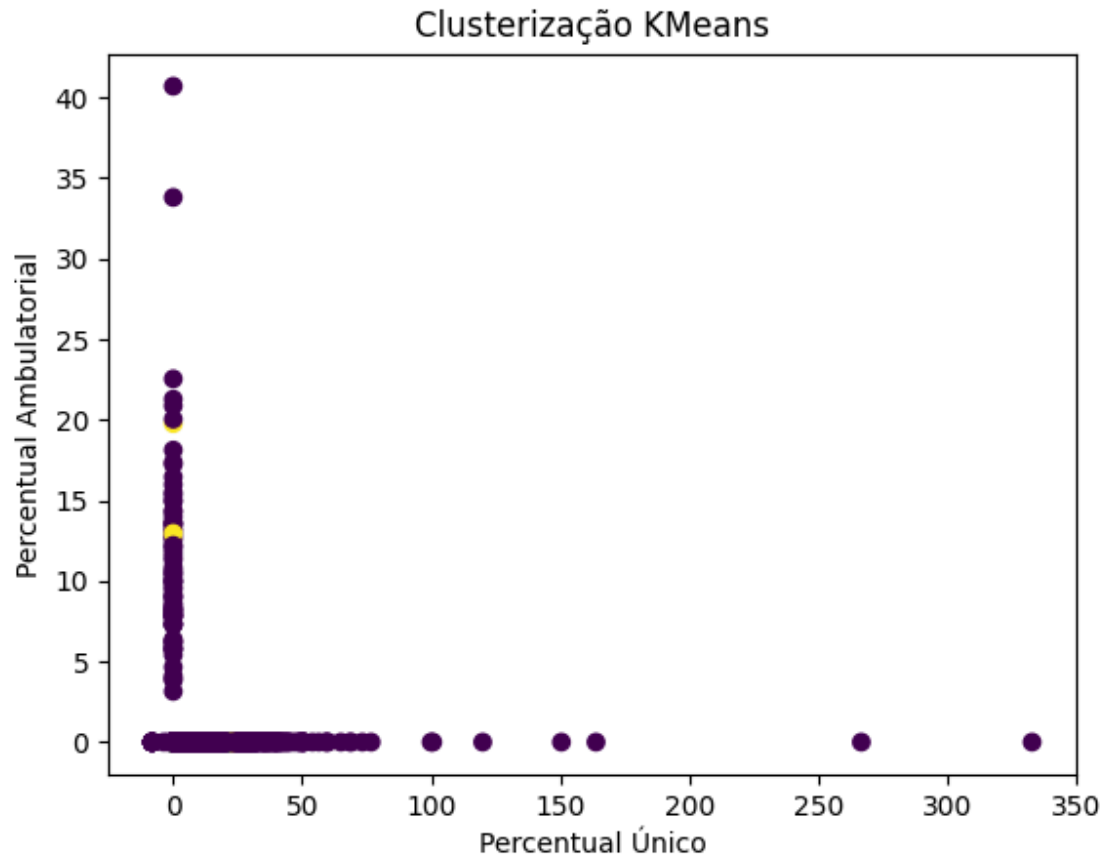
- **Método do Cotovelo:** Esta análise abrange a determinação do número ótimo de clusters, conforme ilustrado na

**[IMAGEM 3] (Método do Cotovelo)**

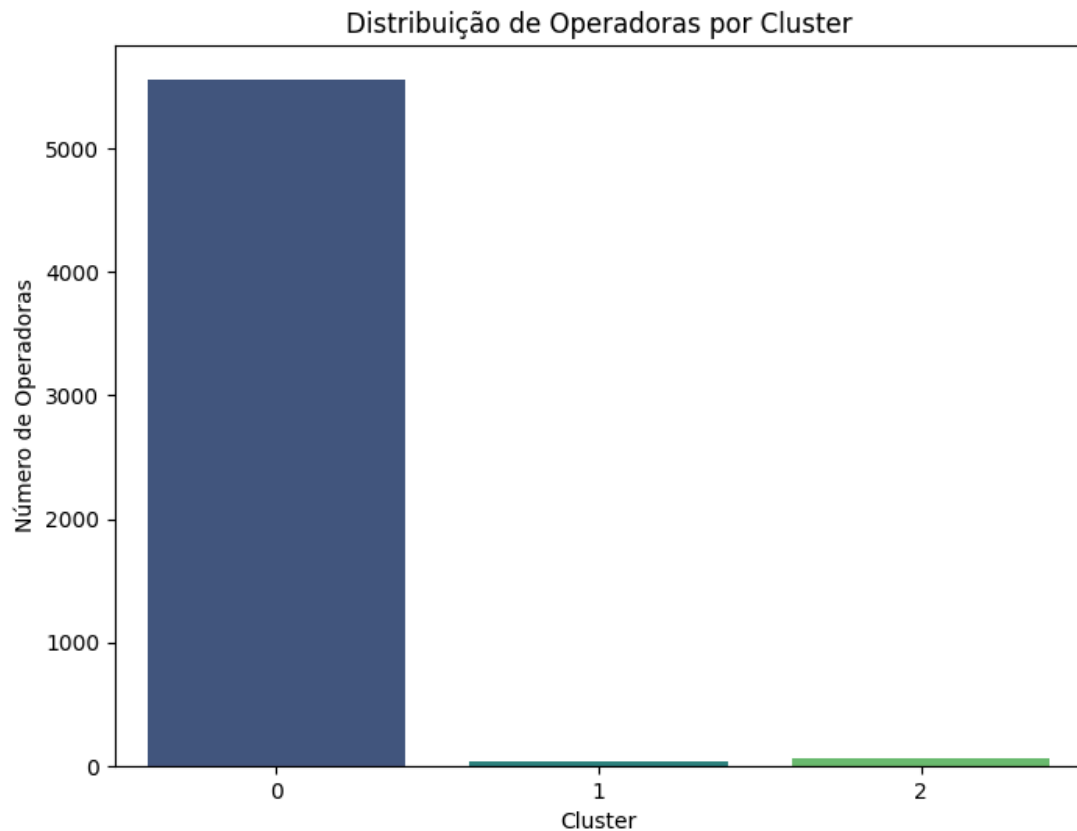


O gráfico indica que o número ideal de clusters é 3, onde a inércia média estabiliza.

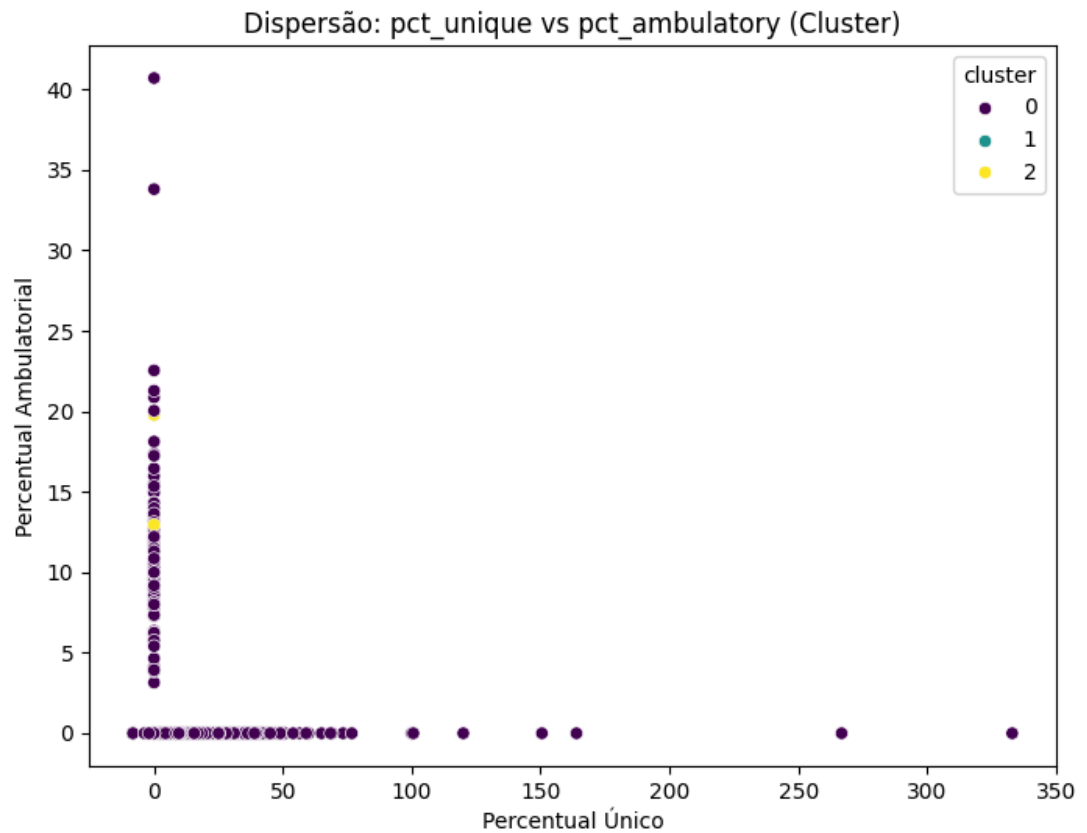
- **Clusterização K-means:** Os resultados da clusterização são apresentados na [IMAGEM 1] (Clusterização K-means)



- A análise mostra a segmentação das operadoras em diferentes grupos com base nos percentuais de reajuste.
- **Distribuição de Operadoras por Cluster:** A maioria (~5.000) está no Cluster 0, como indicado na [IMAGEM 6] (Distribuição de Operadoras por Clusters)



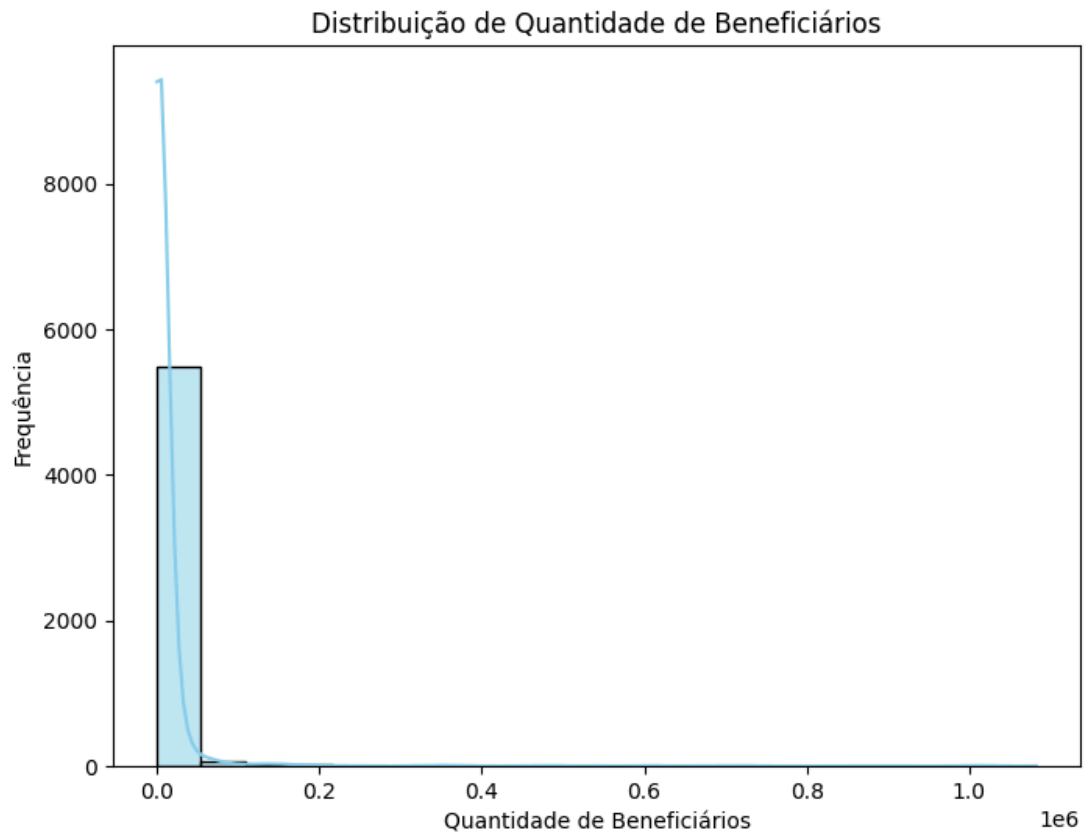
- - 
  - Clusters 1 e 2 têm menos operadoras.
  - **Dispersão (pct\_unique vs. pct\_ambulatory):** A dispersão por clusters é detalhada na
- [IMAGEM 10] (Dispersão: pct\_unique vs. pct\_ambulatory (Cluster))**



- 
- 

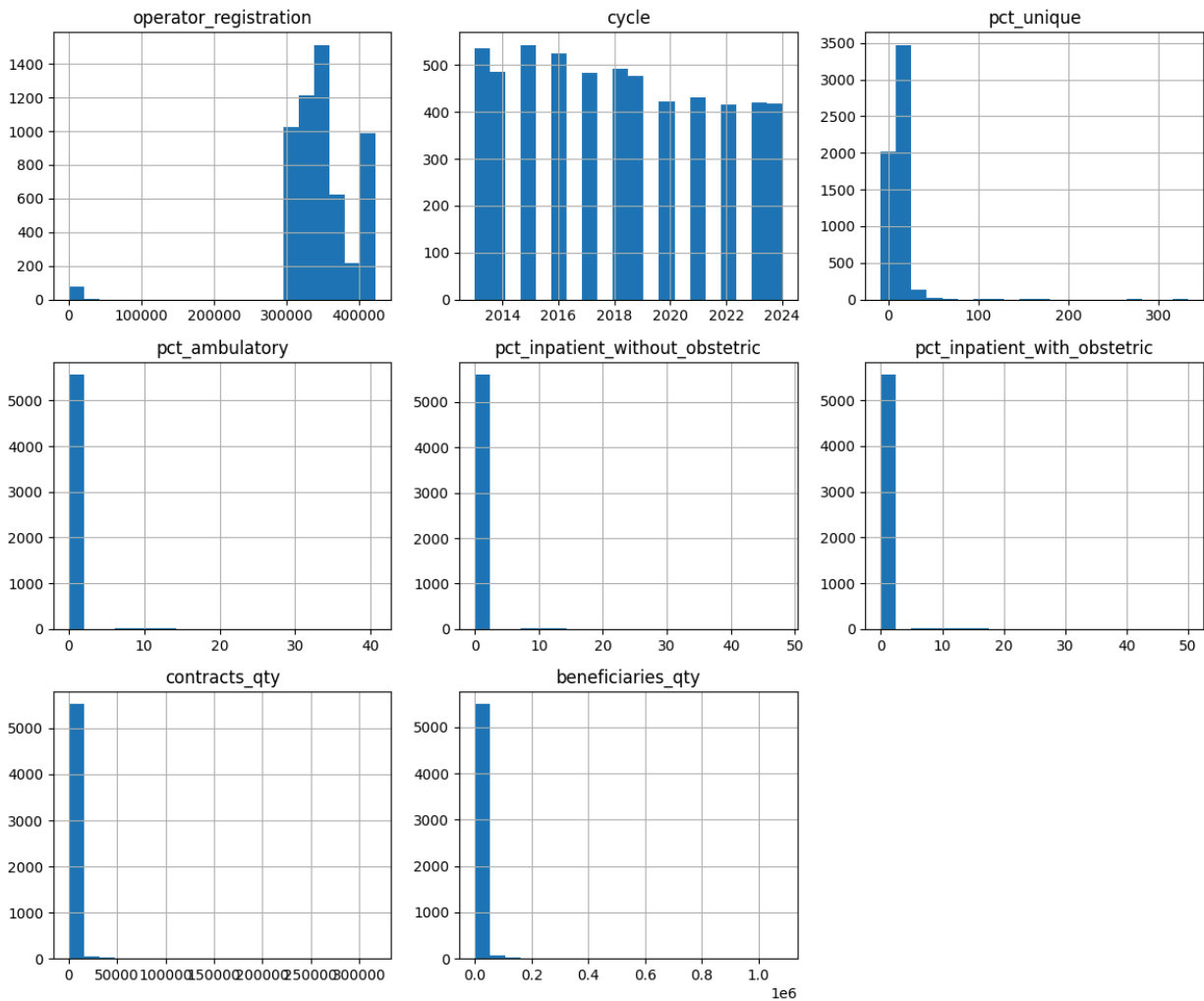
Cluster 0 domina em percentuais ambulatoriais mais altos.

- **Distribuição de Beneficiários:** A concentração de beneficiários está próxima de 0, com pico de ~6.000, conforme a **[IMAGEM 2] (Distribuição de Quantidade de Beneficiários)**

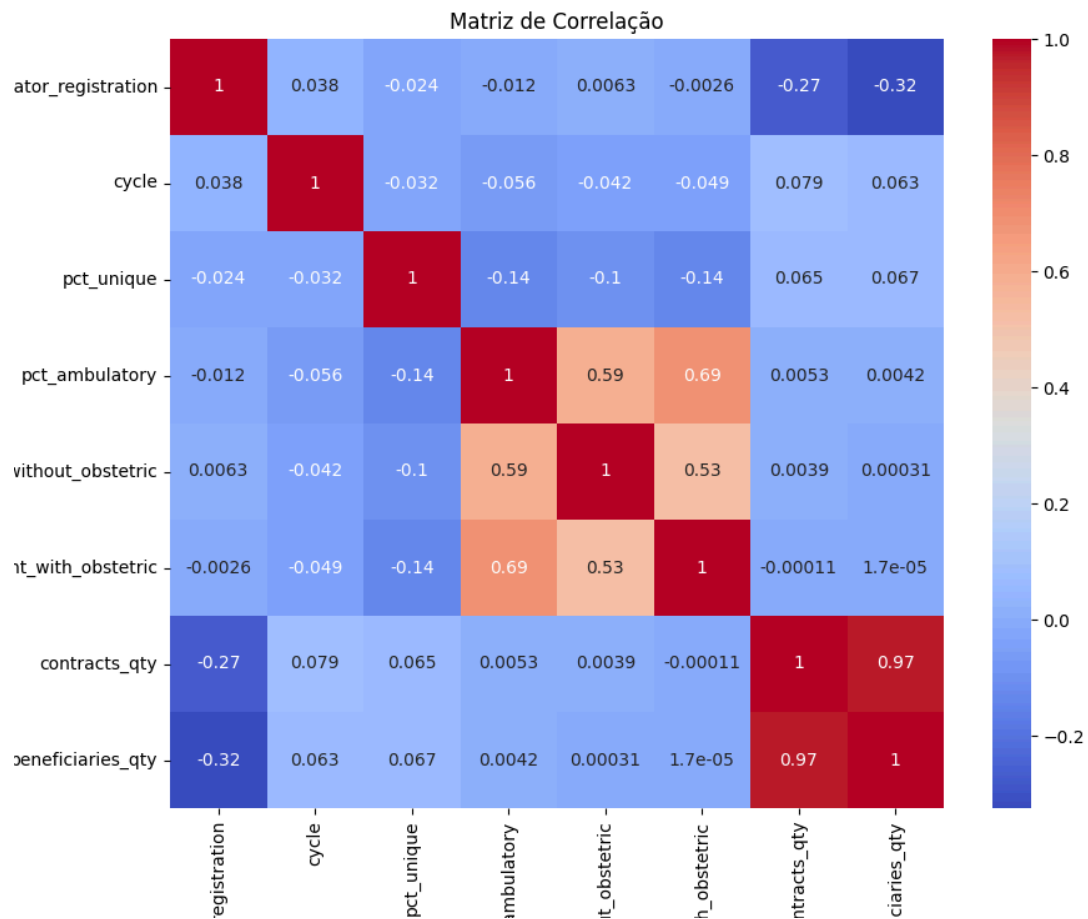


- **Histogramas:** A distribuição de variáveis-chave é explorada nos [IMAGEM 5] (Histogramas)

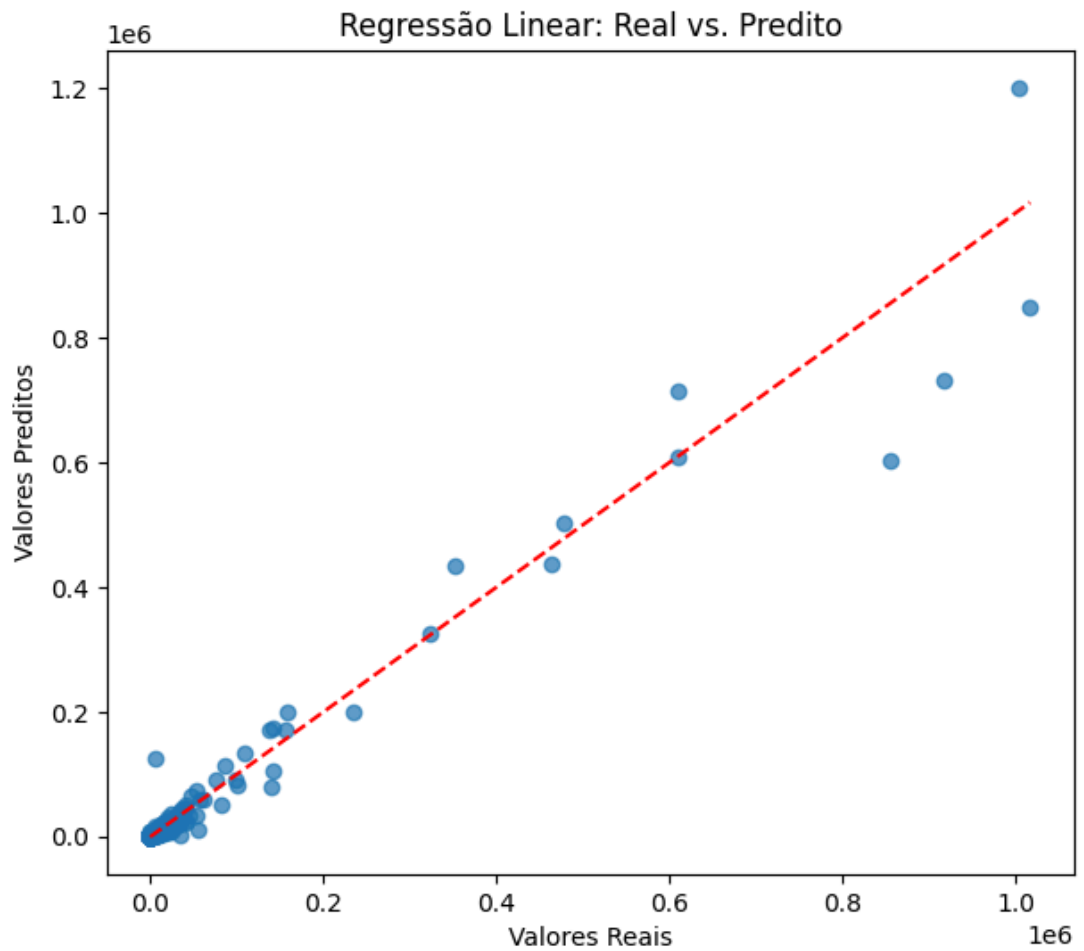




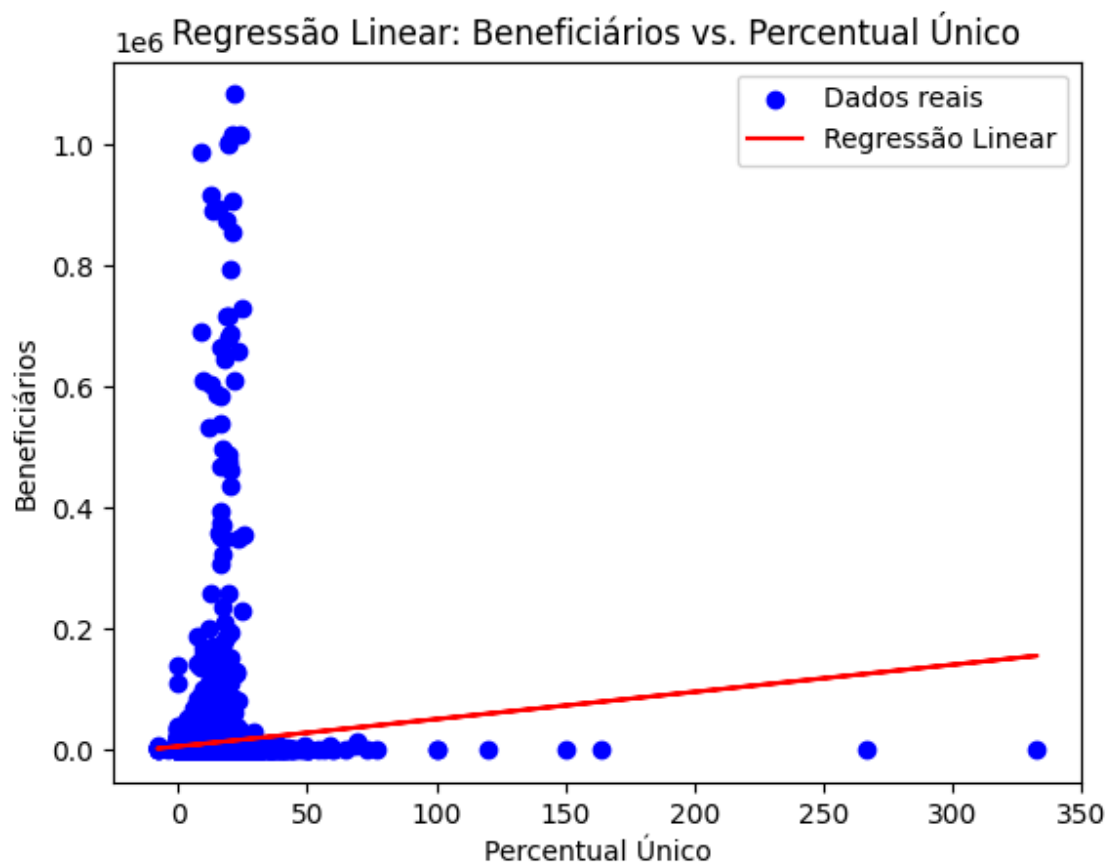
- Esses gráficos ajudam a visualizar a frequência dos percentuais de reajuste e outras métricas.
- **Matriz de Correlação:** As relações entre variáveis como percentual único e número de beneficiários são mostradas na **[IMAGEM 4] (Matriz de Correlação)**



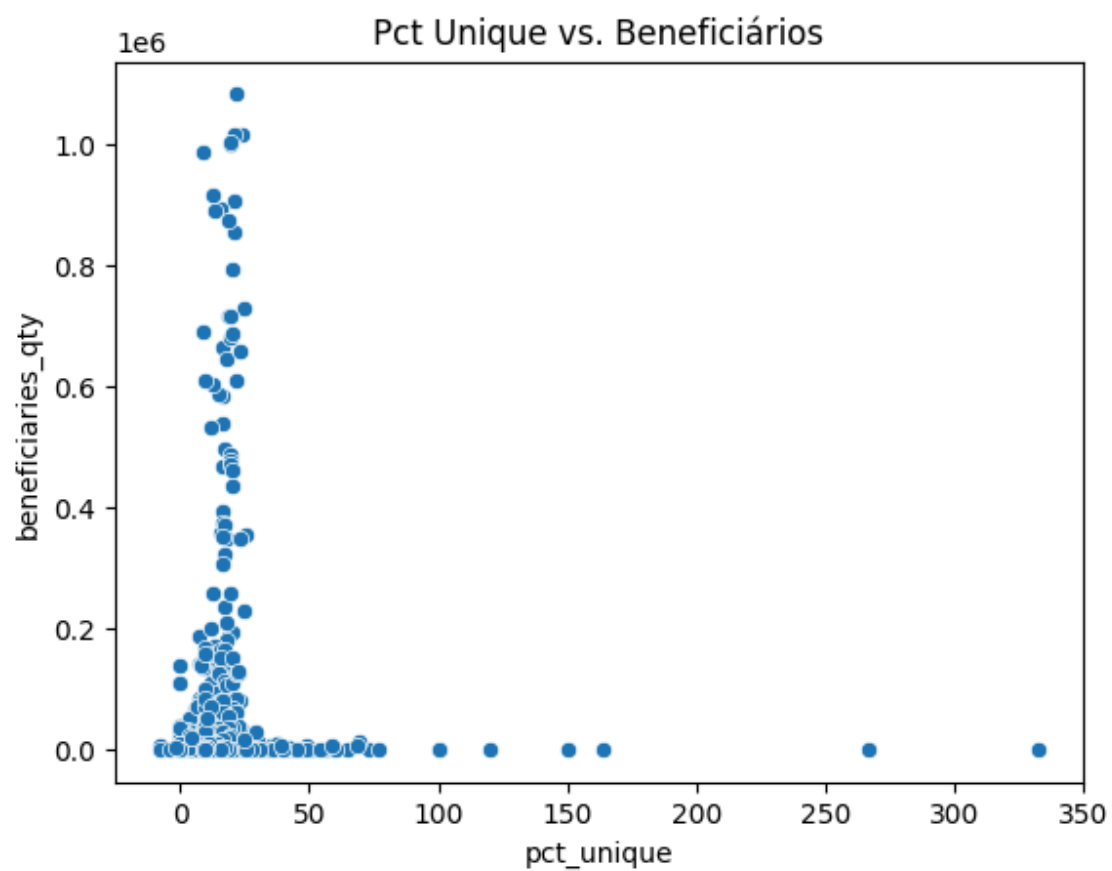
- 
- A matriz destaca correlações fracas entre as variáveis analisadas.
- **Regressão Linear (Real vs. Predito):** A relação entre valores reais e preditos é mostrada na  
**[IMAGEM 7] (Regressão Linear: Real vs. Predito)**



- Os dados apontam uma correlação positiva fraca.
- **Regressão Linear (Beneficiários vs. Percentual Único):** A relação é apresentada na **[IMAGEM 8] (Regressão Linear: Beneficiários vs. Percentual Único)**



- Dados reais mostram baixa variação em percentuais únicos.
- **Percentual Único vs. Beneficiários:** A relação entre essas variáveis é explorada na **[IMAGEM 9] (percentual único e vs. Beneficiários)**

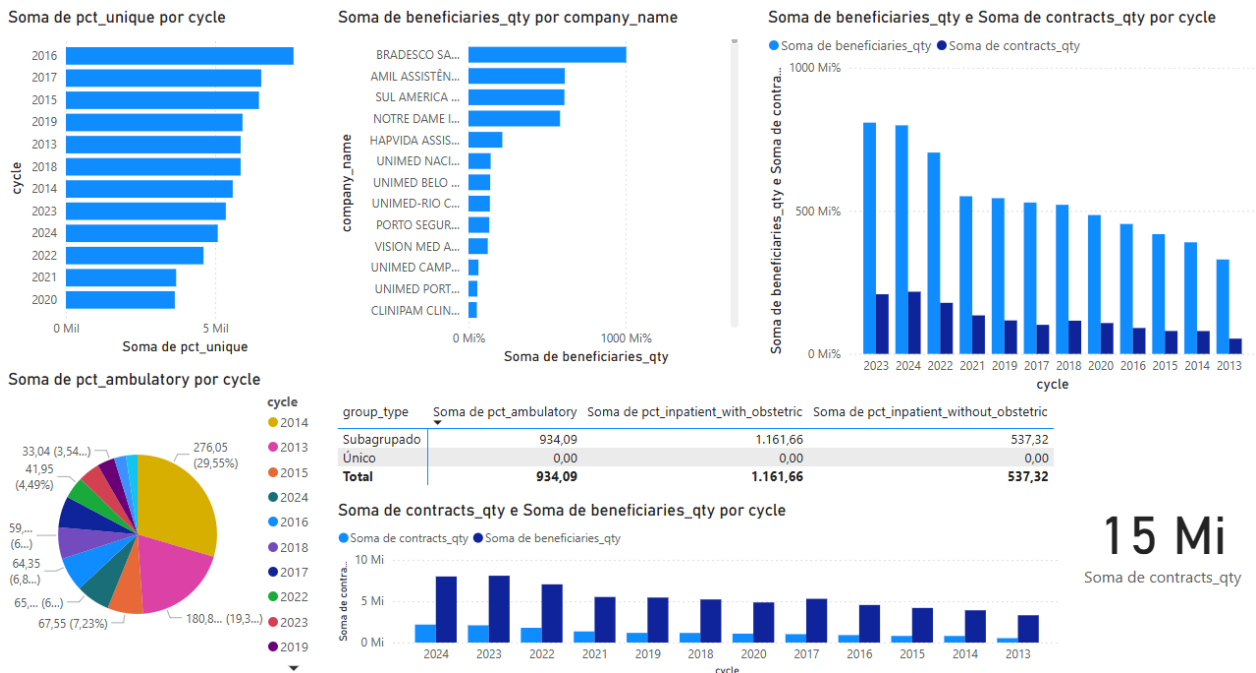


- 

A análise reforça a baixa variação observada na regressão linear.

## 5. Análise dos Gráficos do Power BI

Com o intuito de enriquecer a visualização e interpretação dos dados, gráficos interativos foram desenvolvidos no Power BI com base nos dados da ANS. A seguir, cada visualização é explicada com insights extraídos da análise:



### 1. Gráfico de Barras Horizontais: Soma de pct\_unique por ciclo

- Este gráfico mostra a distribuição do percentual único de reajuste ao longo dos anos.
- Os anos de 2016 e 2017 registram os maiores valores, indicando um pico nos percentuais únicos nesse período, o que pode estar relacionado a mudanças regulatórias ou econômicas.

### 2. Gráfico de Barras: Beneficiários por Operadora

- A maior parte dos beneficiários está concentrada nas grandes operadoras como BRADESCO, AMIL e SULAMÉRICA.
- Essa concentração reforça a importância de políticas específicas para grandes grupos, que afetam uma parcela significativa da população.

### **3. Gráfico de Barras Duplas: Beneficiários e Contratos por Ciclo**

- Visualiza a correlação entre quantidade de beneficiários e de contratos ao longo dos ciclos.
- Os anos de 2023 e 2024 têm os maiores números de beneficiários, porém com queda gradual na quantidade de contratos, sugerindo uma possível consolidação ou migração de beneficiários entre planos.

### **4. Gráfico de Pizza: pct\_ambulatory por ciclo**

- Evidencia que os anos de 2015 e 2014 representam quase 50% do total de percentuais ambulatoriais, sendo os mais relevantes nesse tipo de cobertura.
- Esse destaque temporal pode sinalizar mudanças no perfil de utilização dos planos de saúde ou ajustes na legislação naquele período.

### **5. Tabela Resumo por Tipo de Agrupamento**

- Mostra que os dados analisados pertencem majoritariamente ao grupo "Sub Agrupado", com valores expressivos nos percentuais ambulatoriais e hospitalares com ou sem obstetrícia.

- Isso reforça a representatividade dos agrupamentos utilizados no modelo de clusterização.

## **6. Indicador: Soma Total de Contratos**

- Exibe o total consolidado de contratos na base de dados (15 milhões).
- Esse número corrobora a dimensão e importância do mercado analisado.

## **7. Gráfico de Colunas: contracts\_qty vs. beneficiaries\_qty por ciclo**

- Reforça as observações anteriores, onde o número de beneficiários mantém-se elevado, mas os contratos apresentam variação significativa.
- A diferença entre essas métricas pode refletir planos familiares ou empresariais com múltiplos beneficiários por contrato.

## **6. Análise**

- O método do cotovelo confirma 3 clusters como ideais.
- A regressão linear mostra impacto limitado do percentual único no número de beneficiários.
- A concentração no Cluster 0 sugere uniformidade, enquanto os Clusters 1 e 2 são atípicos.

## **7. Conclusão**



Recomenda-se aprofundar a análise dos Clusters 1 e 2 para identificar fatores diferenciados. O uso de 3 clusters é validado pelo método do cotovelo.

## **8. Importância da Informação**

A análise dos percentuais de reajuste e agrupamento de operadoras de planos de saúde é útil para a comunidade, pois permite identificar padrões que podem impactar diretamente os beneficiários, como diferenças nos custos e acesso a serviços de saúde. Um exemplo direto é o pico de aproximadamente 6.000 beneficiários concentrados em valores próximos de 0 no Cluster 0, o que pode indicar uma grande população dependente de planos com baixo percentual único, possivelmente refletindo uma necessidade de políticas de suporte para evitar aumento de custos ou perda de cobertura. Compreender esses padrões ajuda a promover transparência e pode orientar políticas públicas e decisões da ANS para melhorar a equidade e a qualidade no sistema de saúde suplementar, beneficiando especialmente populações vulneráveis.