

## **Análise do Programa de Subvenção ao Prêmio do Seguro Rural (PSR)**

**Metodologia:** CRISP-DM

**Projeto:** Análise de seguro rural

---

### ***Introdução (Business Understanding)***

O Programa de Subvenção ao Prêmio do Seguro Rural (PSR) foi instituído pelo Governo Federal, por meio do Ministério da Agricultura, Pecuária e Abastecimento (MAPA), com o objetivo de ampliar o acesso dos produtores rurais ao seguro agrícola. O mecanismo central do programa é a concessão de subvenção econômica: parte do valor do prêmio devido pelo produtor à seguradora é financiado pela União. Dessa forma, o custo efetivo do seguro para o produtor é reduzido, incentivando maior adesão e contribuindo para a mitigação de riscos climáticos e de produção.

O SISSER – Sistema de Subvenção Econômica ao Prêmio do Seguro Rural é a base operacional do programa, utilizada na troca de informações entre o MAPA e as seguradoras habilitadas. Nele são registradas as propostas e apólices contempladas com subvenção, contendo dados sobre seguradora, produto SUSEP, cultura/atividade segurada, município/UF da propriedade, área, produtividade estimada, nível de cobertura, prêmio líquido, valor da subvenção, limite de garantia, além de informações de sinistro (indenizações e eventos preponderantes, quando aplicáveis) [【143†dicionariodedados-sisser.pdf】](#) .

Nesta análise, utilizamos o histórico completo de dados disponíveis (2006–2025), que permite uma visão abrangente da evolução do programa. O foco principal está na compreensão do funcionamento e distribuição da subvenção ao longo do tempo, mapeando sua alocação entre seguradoras, culturas agrícolas e regiões do país. Esse mapeamento é essencial para identificar padrões de contratação, concentrações regionais e setoriais, e potenciais assimetrias no acesso ao benefício.

O público-alvo deste artigo são organizações e profissionais do ramo de seguros rurais e inteligência em dados agrícolas, que atuam tanto no apoio a seguradoras e resseguradoras quanto na assessoria a produtores e gestores de risco. A análise aqui conduzida busca fornecer insumos objetivos e padronizados para orientar decisões estratégicas, desenvolvimento de produtos, precificação e monitoramento de riscos.

---

Entendimento dos Dados (Data Understanding)

Fontes e abrangência

O projeto utiliza os dados públicos disponibilizados pelo SISSER, que cobre registros de 2006 a 2025. A granularidade é o nível de proposta/apólice, contemplando desde informações cadastrais do segurado e localização da propriedade até variáveis financeiras e de cobertura. O dicionário de dados oficial foi a principal referência para a interpretação das colunas 【143†dicionariodedados-sisser.pdf】 .

Estrutura dos dados

Os campos disponíveis foram organizados em grupos funcionais:

Tabela 1 – Estrutura dos dados por grupos funcionais

Grupo	Variáveis principais
Identificação	ID_PROPOSTA, NR_PROPOSTA, NR_APOLICE, CD_PROCESSO_SUSEP
Seguradora	NM_RAZAO_SOCIAL
Temporalidade	DT_PROPOSTA, DT_APOLICE, ANO_APOLICE, DT_INICIO_VIGENCIA, DT_FIM_VIGENCIA
Localização	NM_MUNICIPIO_PROPRIEDADE, SG_UF_PROPRIEDADE, CD_GEOCMU, LATITUDE, LONGITUDE, NR_GRAU_LAT, NR_MIN_LAT, NR_SEG_LAT, NR_GRAU_LONG, NR_MIN_LONG, NR_SEG_LONG
Objeto segurado	NM_CULTURA_GLOBAL, NM_CLASSIF_PRODUTO, NR_AREA_TOTAL, NR_ANIMAL
Produtividade	NR_PRODUTIVIDADE_ESTIMADA, NR_PRODUTIVIDADE_SEGURADA
Cobertura e valores	NivelDeCobertura, VL_LIMITE_GARANTIA, VL_PREMIO_LIQUIDO, PE_TAXA, VL_SUBVENCAO_FEDERAL
Sinistros	VALOR_INDENIZAÇÃO, EVENTO_PREPONDERANTE

## **Qualidade dos dados**

- Inconsistências temporais em datas de proposta e vigência.
- Campos nulos ou não aplicáveis (nr\_animal em culturas vegetais).
- Outliers financeiros e de produtividade.
- Redundâncias em nm\_cultura\_global e evento\_preponderante.
- Duplicidades potenciais em propostas/apólices.

Para análises detalhadas sobre a quantidade de flags de qualidade, ver Notebook de Análises Exploratórias – *Tabela 2*.

---

## ***Preparação dos Dados (Data Preparation)***

### **Estrutura em camadas**

- Raw: arquivos originais.
- Interim: base padronizada, limpa e categorizada.
- Processed: base final em Parquet particionado, consolidada com flags, métricas derivadas e chaves de segurados.

### **Regras aplicadas**

- Flags de qualidade para prêmios, garantias, produtividades e datas.
- Métricas derivadas: taxa de prêmio, sinistralidade, subvenção relativa, área média segurada, variáveis temporais.
- Identificação de segurados por chave única.
- Padronização de tipos (strings, datetime, floats).
- Testes de integridade estrutural e semântica.
- Flags adicionais para monitorar prêmios nulos/negativos e garantias inválidas.

Para visualizar o quantitativo de flags, consultar Notebook de Análises Exploratórias – *Tabela 2*.

---

## **Modelagem e Resultados (Exploração de Dados)**

### **Métricas analisadas**

O dataset processado possibilitou o cálculo de métricas relevantes para avaliação da política pública e do mercado de seguros rurais:

- **Taxa de prêmio (METR\_TAXA\_PREMIO):** proporção entre prêmio líquido e valor segurado.
- **Sinistralidade (METR\_SINISTRALIDADE):** razão entre indenizações e prêmio líquido.
- **Área média segurada por apólice (METR\_AREA\_MEDIA):** dimensão média das áreas cobertas.
- **Subvenção relativa (METR\_SUBVENCAO\_RELATIVA):** participação da subvenção federal no prêmio.

### **Concentração regional e setorial**

As análises indicaram forte **concentração geográfica**:

- O Paraná lidera em número de apólices, seguido por Minas Gerais e Rio Grande do Sul.
- Esse padrão mostra que poucos estados absorvem a maior parte da política, enquanto outros, com base agrícola relevante, apresentam baixa participação. (Ver Notebook – Tabela 4 e Gráfico 5).

Do ponto de vista setorial, culturas específicas concentram a maior parte dos contratos e da subvenção, reforçando o caráter focalizado do programa.

### **Eficiência da subvenção**

A **eficiência média** do programa (quanto de valor segurado cada R\$ 1 de subvenção gera) apresenta grandes variações:

- Enquanto alguns anos e culturas mostram valores superiores a R\$ 100 por R\$ 1 de subvenção, outros ficam abaixo de R\$ 50.
- Isso sugere que a alocação da subvenção nem sempre maximiza o valor segurado, levantando oportunidade de revisão da política. (Ver Notebook – Tabela 3).

### **Risco e sinistralidade**

A sinistralidade média variou bastante no período analisado:

- Patamares mais altos em determinados anos e culturas indicam maior exposição ao risco.

- Há evidências de **concentração de indenizações em poucos estados**, o que pode afetar o equilíbrio atuarial do programa. (Ver Notebook – Tabela 3 e Gráfico 4).

### Evolução histórica

- Entre 2006 e 2010 (recorte do relatório exploratório), foram registradas mais de 230 mil apólices, cobrindo cerca de 19 milhões de hectares.
- O crescimento não é linear: houve anos de forte expansão (2008–2009) seguidos de redução em 2010.
- Essa oscilação pode estar relacionada à disponibilidade orçamentária da União ou a fatores climáticos. (Ver Notebook – Tabela 3 e Gráfico 2).

### Oportunidades estratégicas

- Estados com grande área agrícola mas baixa participação no PSR representam **mercados potenciais para expansão**.
- Culturas menos contempladas podem ser alvo de políticas de incentivo, diversificando o risco e ampliando a resiliência do setor.
- A chave única de segurados permite identificar **clientes recorrentes e fiéis**, possibilitando análises de fidelização e estratégias comerciais específicas.

---

### Discussão e Limitações

- **Lacunas de dados:** elevada proporção de nulos em variáveis financeiras e de produtividade reduz a confiabilidade de análises de eficiência.
- **Cobertura temporal:** embora a base cubra 2006–2025, a completude varia entre anos.
- **Viés de concentração:** resultados refletem culturas e regiões de maior participação no PSR, podendo subestimar pequenas culturas/UFs.
- **Dados pessoais:** CPF/CNPJ foram normalizados e tratados apenas como chaves analíticas, respeitando boas práticas de governança e LGPD.

---

### Conclusão

A análise do PSR evidencia avanços na estruturação e disponibilização dos dados, permitindo compreender padrões de distribuição da subvenção e perfis de contratação. O pipeline desenvolvido garantiu a criação de uma base **curada, padronizada e enriquecida**, apta para análises estatísticas e modelos preditivos.

As métricas derivadas e flags de qualidade ampliaram a capacidade de monitorar eficiência e riscos.

Próximos passos incluem:

- Exploração de dashboards interativos.
- Modelagem preditiva de risco de sinistro.
- Avaliação longitudinal de segurados para estudos de fidelização.
- Integração com dados externos (clima, produção agrícola) para análises causais.