

# **Análise Preditiva no Agronegócio: Relação entre Safra e Preço do Café no Brasil**

**Alana Cristina Martens**  
**Camilli Frigeri Feliz**  
**Samantha de Souza Andrade**  
**Mariele Barbosa da Silva**

---

## ***Resumo do Projeto***

O presente trabalho tem como objetivo analisar como a variação da produção de café (safra) se relaciona com o preço do café arábica no Brasil. Parte-se da hipótese de que a redução na produção tende a elevar o preço médio da saca. Dessa forma, o estudo busca compreender a dinâmica entre oferta e valor de mercado, utilizando métodos estatísticos e práticas de governança de tecnologia e dados.

---

## **1. Governança Corporativa e Governança de TI aplicadas ao projeto**

A governança corporativa estabelece princípios de direção, controle e prestação de contas voltados para a geração de valor sustentável. Já a Governança de TI (GTI) assegura que os recursos tecnológicos apoiem a estratégia organizacional, promovendo a gestão de riscos, conformidade legal e a entrega de benefícios para o negócio.

### **Princípios aplicados**

- Transparência, equidade, responsabilidade e accountability.
- Na TI, esses princípios se traduzem em alinhamento estratégico, entrega de valor, gerenciamento de riscos, otimização de recursos e mensuração de desempenho.

### **Estrutura de papéis**

- Conselho ou patrocinador: define objetivos e apetite de risco.
- Comitê de Dados e TI: prioriza iniciativas.
- Product Owner de Analytics: organiza o backlog.
- Equipe de Dados e Machine Learning: implementa soluções.
- Jurídico e Segurança da Informação: asseguram conformidade.
- Auditoria interna: revisa controles.

## Processos e frameworks de apoio

- **COBIT:** objetivos de controle e desempenho.
- **ITIL:** gestão de serviços de TI.
- **ISO 27001 e 27701:** segurança da informação e privacidade.
- **ISO 31000:** gestão de riscos.
- **CRISP-DM e MLOps:** ciclo de vida de modelos preditivos.

## Arquitetura tecnológica necessária

- **Hardware:** servidores ou nuvem com CPU/GPU, armazenamento seguro e backups criptografados.
- **Software:** Python/R, notebooks, bibliotecas de Machine Learning, orquestradores (Airflow/Prefect), controle de versão (Git) e ferramentas de qualidade de dados.
- **Rede:** segmentação de ambientes, VPN, firewall, IDS/IPS e políticas de acesso Zero Trust.
- **Banco de Dados:** Data Lake/Lakehouse para dados brutos e processados, criptografia em repouso (AES-256) e em trânsito (TLS).
- **Coleta de dados:** conectores para CEPEA, CONAB e IBGE, ETL/ELT com auditoria e validações automáticas.

## Gestão de riscos

Inclui monitoramento de qualidade dos dados, mitigação de viés nos modelos, prevenção de vazamentos de informação, contingência contra indisponibilidade de sistemas e acompanhamento da conformidade regulatória.

---

## 2. LGPD aplicada ao projeto

A **Lei Geral de Proteção de Dados (LGPD – Lei nº 13.709/2018)** define diretrizes para o tratamento adequado de informações pessoais, reforçando princípios éticos e de segurança.

### Aplicações no projeto

- **Bases legais:** preferencialmente anonimização. Caso sejam usados dados pessoais (cadastros de fornecedores ou produtores), aplicar consentimento, obrigação legal ou legítimo interesse.
- **Princípios:** finalidade, adequação, necessidade, qualidade, transparência, segurança, prevenção e responsabilização.
- **Segurança por design:** criptografia, pseudonimização, segregação de ambientes, autenticação multifator e registros de auditoria.
- **Direitos dos titulares:** mecanismos para solicitação de acesso, correção, exclusão e portabilidade dos dados.

- **Gestão de terceiros:** contratos com operadores contendo cláusulas de proteção de dados.
  - **Resposta a incidentes:** plano com prazos, comunicação à ANPD e registro das ações tomadas.
  - **Governança:** indicação de encarregado (DPO), treinamentos, auditorias periódicas e relatórios de conformidade.
- 

### 3. Resumo de Estatística Descritiva

Considerando uma amostra de  $n$  observações  $(x_1, x_2, \dots, x_n)$ :

- **Média amostral:**  

$$\bar{x} = (1/n) \cdot \sum x_i$$
- **Mediana:** valor central após ordenar os dados.
- **Moda:** valor mais frequente (pode haver mais de um).
- **Variância amostral:**  

$$s^2 = \sum (x_i - \bar{x})^2 / (n - 1)$$
- **Desvio-padrão amostral:**  

$$s = \sqrt{s^2}$$
- **Variância populacional:**  

$$\sigma^2 = \sum (x_i - \mu)^2 / N$$
- **Desvio-padrão populacional:**  

$$\sigma = \sqrt{\sigma^2}$$
- **Covariância amostral entre X e Y:**  

$$s_{xy} = \sum [(x_i - \bar{x})(y_i - \bar{y})] / (n - 1)$$

#### Interpretação

- Média, mediana e moda: medidas de tendência central.
- Variância e desvio-padrão: grau de dispersão dos dados.
- Covariância: indica a direção da relação entre duas variáveis (positiva ou negativa).
- Aplicação no projeto: análise da relação entre **produção (safra)** e **preço do café**, calculando correlação para verificar dependência entre variáveis.