# Universidade Positivo

Francisco Bueno Ghizelini  
 Guilherme Candida

Análise de Dados do Projeto *Preço Justo PR*

Curitiba, Paraná

2025

Francisco Bueno Ghizelini  
 Guilherme Candida

Análise de Dados do Projeto *Preço Justo PR*

Relatório do trabalho A2 da disciplina de

Data science, apresentada pelo professor Rubem Matimoto Koide pela Universidade Positivo

Curitiba, Paraná

2025

# 

## **1. Governança Corporativa e Governança de TI**

# Para contextualizar a aplicação da governança, definimos nosso projeto como a criação de uma empresa de tecnologia denominada *Preço Justo PR*. O negócio consiste em uma plataforma digital (website e aplicativo) que coleta, analisa e apresenta os preços dos combustíveis no estado do Paraná, com o objetivo de promover transparência ao consumidor.

### **1.1 Governança Corporativa**

# A Governança Corporativa é o sistema que dirige e controla a empresa *Preço Justo PR*, garantindo que os interesses de sócios, clientes e da sociedade sejam atendidos de forma ética e sustentável. A aplicação se dá com base nos quatro pilares:

# **Transparência**: a plataforma informará de forma explícita que os dados são provenientes da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). A última atualização será sempre visível, bem como a metodologia de cálculo do “preço médio” ou “economia potencial”.

# **Equidade**: todos os postos cadastrados na base da ANP serão tratados de forma isonômica, sem favorecimentos ou destaques patrocinados. O mesmo princípio valerá para os usuários, com acesso equitativo às funcionalidades.

# **Prestação de Contas**: os gestores deverão prestar contas a investidores e usuários quanto ao desempenho da plataforma, a veracidade dos dados e as decisões estratégicas. Em caso de erros, a responsabilidade será assumida, com comunicação e correção imediatas.

# **Responsabilidade Corporativa**: a empresa zelará pela sustentabilidade do negócio e pelo impacto social positivo, respeitando a proteção de dados dos usuários (conforme a LGPD) e contribuindo para um mercado de combustíveis mais justo.

### **1.2 Governança de TI**

# A Governança de TI será o coração do negócio, alinhada aos princípios da Governança Corporativa, com a função de garantir suporte tecnológico às metas estratégicas.

# **Banco de Dados**: uso de um SGBD relacional robusto (PostgreSQL) com rotinas automatizadas de ETL para atualização diária dos dados da ANP.

# **Infraestrutura**: hospedagem em nuvem (AWS, Google Cloud ou Azure), com escalabilidade automática e balanceamento de carga, além de CDN para garantir velocidade de acesso.

# **Segurança**: práticas de desenvolvimento seguro (OWASP Top 10), criptografia de dados em trânsito e em repouso, firewall de aplicação e autenticação reforçada.

# **Arquitetura**: utilização de microsserviços para permitir escalabilidade modular (busca de preços, login, relatórios).

# **Business Intelligence**: ferramentas de BI e Data Analytics para monitoramento do mercado e apoio às decisões estratégicas.

# 

## **2. Aplicação da Lei Geral de Proteção de Dados (LGPD)**

# A LGPD (Lei nº 13.709/2018) será aplicada de forma central no *Preço Justo PR*, garantindo confiança e proteção ao usuário.

### **2.1 Dados Tratados**

# **Dados Públicos de Empresas**: CNPJ, razão social e endereço do posto, extraídos do CSV da ANP.

# **Dados Pessoais de Usuários**: nome, e-mail e senha para cadastro; localização geográfica (GPS) somente quando autorizado para busca de postos próximos.

### **2.2 Relação com Ética, Segurança e Proteção**

# **Finalidade**: dados de localização utilizados apenas para encontrar postos próximos. Não haverá rastreamento ou marketing não autorizado.

# **Segurança**: senhas com hashing forte (bcrypt), criptografia ponta a ponta e auditorias periódicas.

# **Transparência**: painel de privacidade para visualizar, corrigir ou excluir dados pessoais. Política de privacidade em linguagem clara.

# **Autenticação**: login seguro com 2FA para administradores.

# **Minimização de Dados**: coleta apenas do necessário, sem exigir CPF ou telefone.

# 

## **3. Resumo dos Conceitos de Estatística Descritiva**

| **Conceito** | **Definição** | **Equação Amostral** |
| --- | --- | --- |
| Média | Soma de todos os valores dividida pelo número de observações. | x̄ = Σxi / n |
| Mediana | Valor central em um conjunto ordenado. | (n+1)/2 |
| Moda | Valor mais frequente em um conjunto. | – |
| Variância | Dispersão dos dados em relação à média. | s² = Σ(xi – x̄)² / (n–1) |
| Desvio Padrão | Raiz quadrada da variância, indicando afastamento médio em relação à média. | s = √s² |
| Covariância | Mede a relação linear entre duas variáveis. | Cov(X,Y) = Σ(xi – x̄)(yi – ȳ) / (n–1) |
| Assimetria | Mede o grau e a direção da inclinação da distribuição em relação à média. | g1 = Σ(xi – x̄)³ / [(n–1)·s³] |
| Curtose | Mede o grau de “achatamento” ou “concentração” da distribuição em torno da média. | g2 = Σ(xi – x̄)⁴ / [(n–1)·s⁴] |

# 

## **4. Cálculo Comparativo e Análise dos Resultados**

# A análise foi realizada sobre os dados do arquivo *Preços\_Automotivos\_Parana.xlsx*.

### **4.1 Estatística Descritiva**

# (Insira aqui as tabelas de resultados já processadas no Python).

### **4.2 Matriz de Covariância**

# (Insira aqui a matriz gerada no Python).

### **4.3 Gráficos**

# **Histograma**: mostra a distribuição dos preços da Gasolina Comum.

# **Box Plot**: compara a dispersão dos preços entre combustíveis.

# **Gráfico de Percentis**: mostra visualmente a posição dos quartis e percentis (25%, 50%, 75%), permitindo identificar a dispersão e concentração dos preços.

### **4.4 Interpretação**

# A Gasolina Aditivada apresentou maior média (R$ 6,14).

# O Etanol foi o combustível mais volátil (maior desvio padrão).

# A Gasolina Comum apresentou menor dispersão, sugerindo preços mais homogêneos.

# A covariância positiva confirma que os preços variam em conjunto.

# A análise de assimetria e curtose mostrou distribuições com leve inclinação e concentração diferenciada dos preços, confirmando a existência de outliers.

# 

## **5. Instruções para Cálculo em Planilhas**

# Exemplo de fórmulas no Excel ou Google Sheets (supondo preços na coluna M):

# Média: =AVERAGE(M2:M1000)

# Mediana: =MEDIAN(M2:M1000)

# Moda: =MODE.SNGL(M2:M1000)

# Desvio Padrão: =STDEV.S(M2:M1000)

# Variância: =VAR.S(M2:M1000)

# Percentil 25: =PERCENTILE.INC(M2:M1000;0,25)

# Covariância: =COVARIANCE.S(intervalo1; intervalo2)

# 

## **6. Considerações Finais**

# O projeto *Preço Justo PR* mostra-se viável tanto do ponto de vista tecnológico quanto de análise de dados. A aplicação da Governança Corporativa e da Governança de TI garante sustentabilidade e alinhamento estratégico, enquanto a LGPD assegura proteção dos dados.

# As análises estatísticas reforçam a importância do monitoramento contínuo dos preços e evidenciam tendências de mercado que impactam diretamente o consumidor.

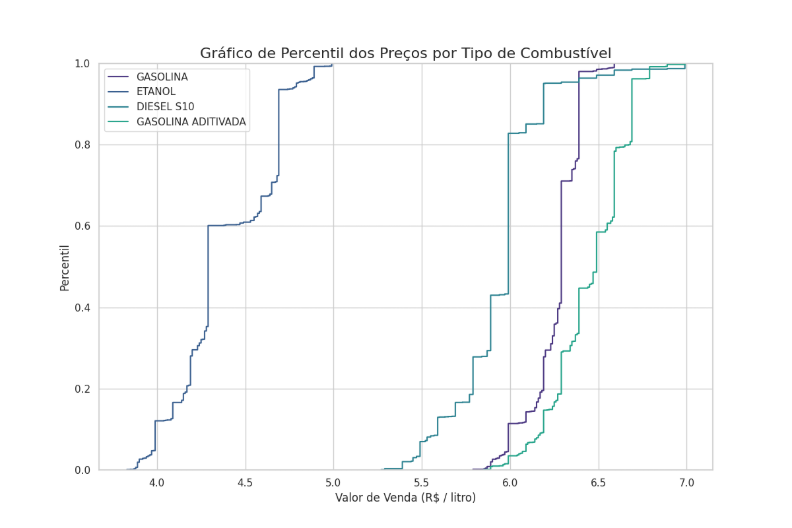
# 

## **Referências**

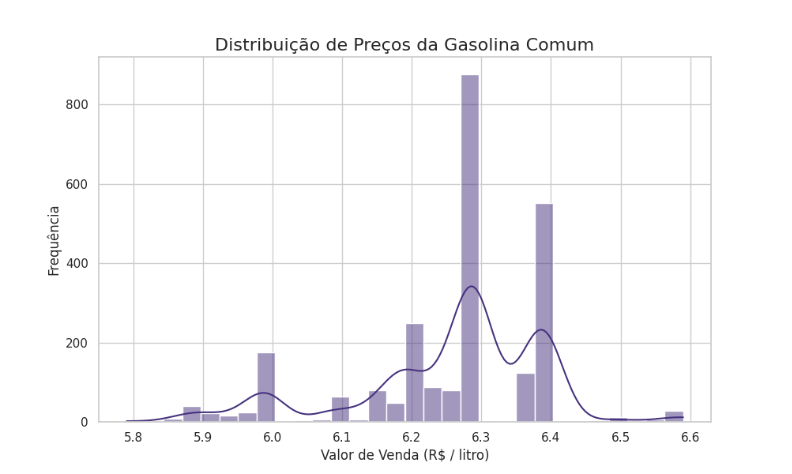
# AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS (ANP). Dados de preços de combustíveis. Disponível em: https://www.gov.br/anp. Acesso em: 05 set. 2025.

# BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: http://www.planalto.gov.br. Acesso em: 05 set. 2025.

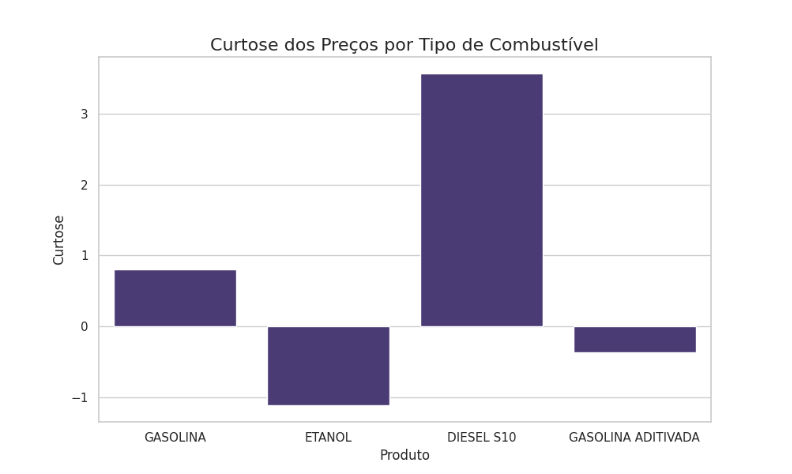
**Gráficos Python**

****

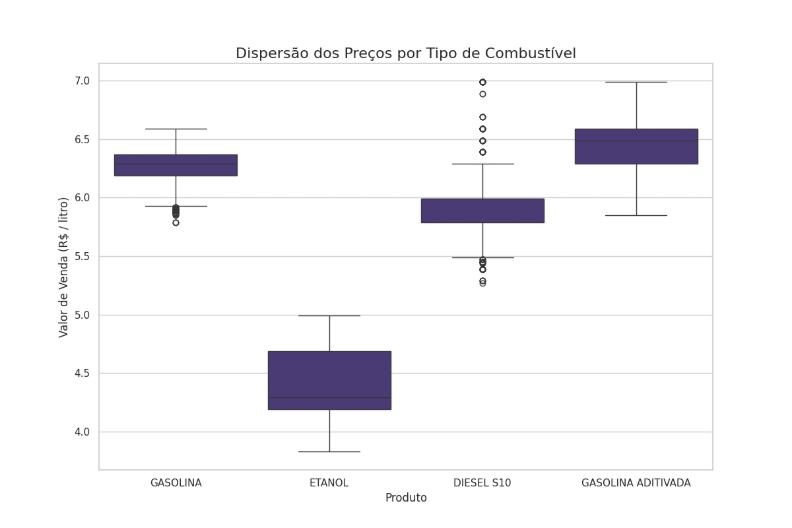
**Gráfico de Percentil dos Preços por Tipo de Combustível**

****

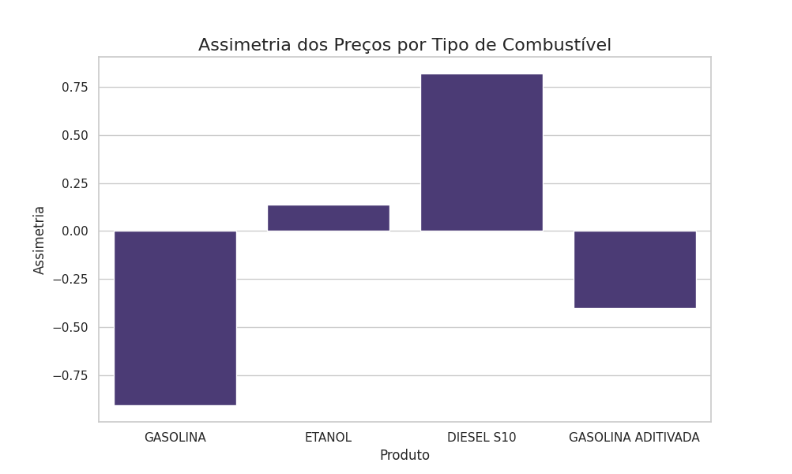
**Distribuição de Preços da Gasolina Comum**

****

**Curtose dos Preços por Tipo de Combustível**

****

**Dispersão dos Preços por Tipo de Combustível**

****

**Assimetria dos Preços por Tipo de Combustível**