

## **Challenge Sprint 2**

**Coleta Automatizada de Dados  
de Produtos HP em  
E-commerce (RPA)**

## Conteúdo

<b>Fontes Adicionais Utilizadas</b>	<b>2</b>
<b>Estratégia de Extração e Integração Adotada</b>	<b>2</b>
<b>Desafios Encontrados e Soluções Aplicadas</b>	<b>2</b>
<b>Etapas de Padronização e Fusão com a Base Principal</b>	<b>2</b>
<b>Processo de Tratamento dos Dados</b>	<b>2</b>
<b>Criação das Colunas Derivadas</b>	<b>2</b>
<b>Estratégia de Coleta de Novas Fontes e Justificativa</b>	<b>3</b>
<b>Análise Descritiva e Visual dos Dados</b>	<b>3</b>
Distribuição de Preços	3
Preço Mercado Livre vs Preço Real por Produto	4
Distribuição de Preço do Mercado Livre por Produto Real	4

## Informações do Documento

<b>Projeto:</b>	Challenge Sprint 2025.2 – RPA – 1º semestre		
<b>Equipe:</b>	555677 - Matheus Hungaro Fidelis 556389 - Pablo Menezes Barreto 556984 - Tiago Toshio Kumagai Gibo 554668 - Israel Dalcin 555183 - Danilo Ramalho Silva 556213 - João Vitor Pires da Silva	<b>Versão do Documento:</b>	1.0

## Histórico de versão

<b>Versão</b>	<b>Data</b>	<b>Revisado por</b>	<b>Descrição</b>	<b>Nome do Arquivo</b>
1.0	15/06	Tiago Toshio Kumagai Gibo	Criação do documento	Challenge Sprint 2025.2 - RPA

# Relatório Técnico de Integração de Dados

---

## Fontes Adicionais Utilizadas

- Lista de preços fornecida pela HP (arquivo CSV).
- Dados extraídos do site Mercado Livre com Selenium (banco SQLite em mercadolive.db).

## Estratégia de Extração e Integração Adotada

- Com nossa base de dados e a lista de preços, utilizamos o cruzamento das informações junto com a média de análise de sentimentos dos comentários

## Desafios Encontrados e Soluções Aplicadas

- Conteúdo carregado dinamicamente exigiu waits explícitos no Selenium.
- Estrutura variável de página tratada com normalização de colunas e tratamento de missing values.
- Timeouts resolvidos com retries e ajuste de parâmetros.

## Etapas de Padronização e Fusão com a Base Principal

- Normalização de nomes de colunas (lowercase e underscores).
- Conversão de tipos de dados (ex: strings de preço para float).
- Integração conforme estratégia definida.

## Processo de Tratamento dos Dados

- Remoção de duplicatas.
- Imputação simples para valores ausentes.
- Conversão de datas para datetime em análises futuras.

## Criação das Colunas Derivadas

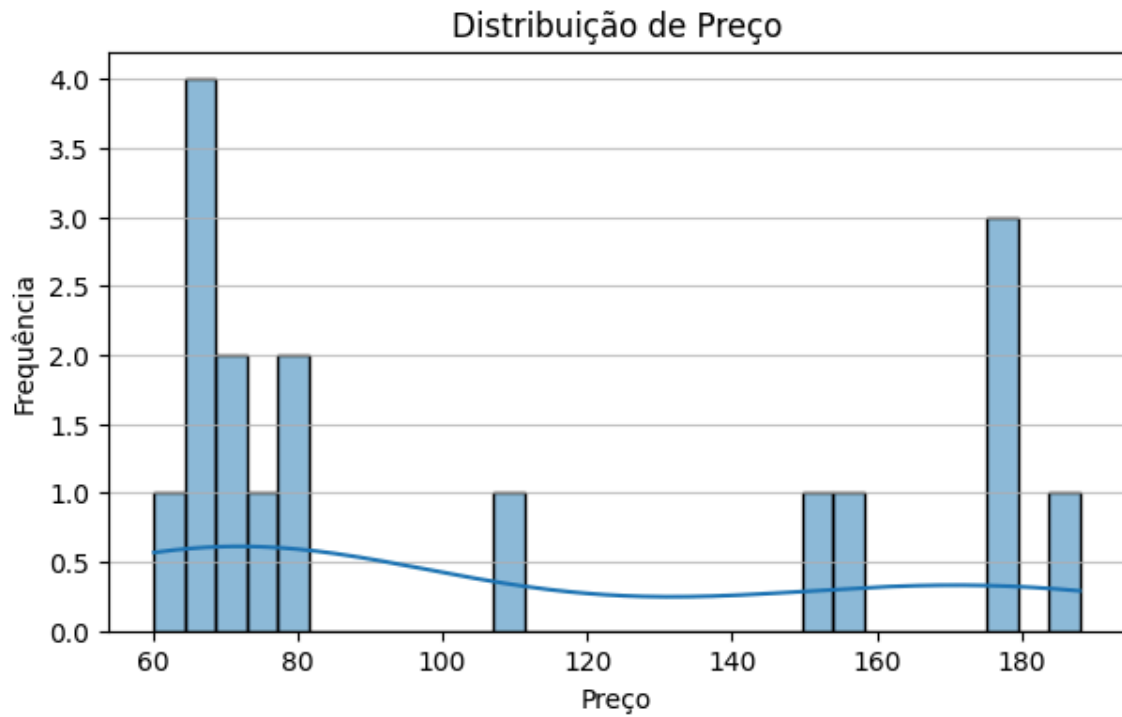
- mean\_label, real\_price e real\_product

## Estratégia de Coleta de Novas Fontes e Justificativa

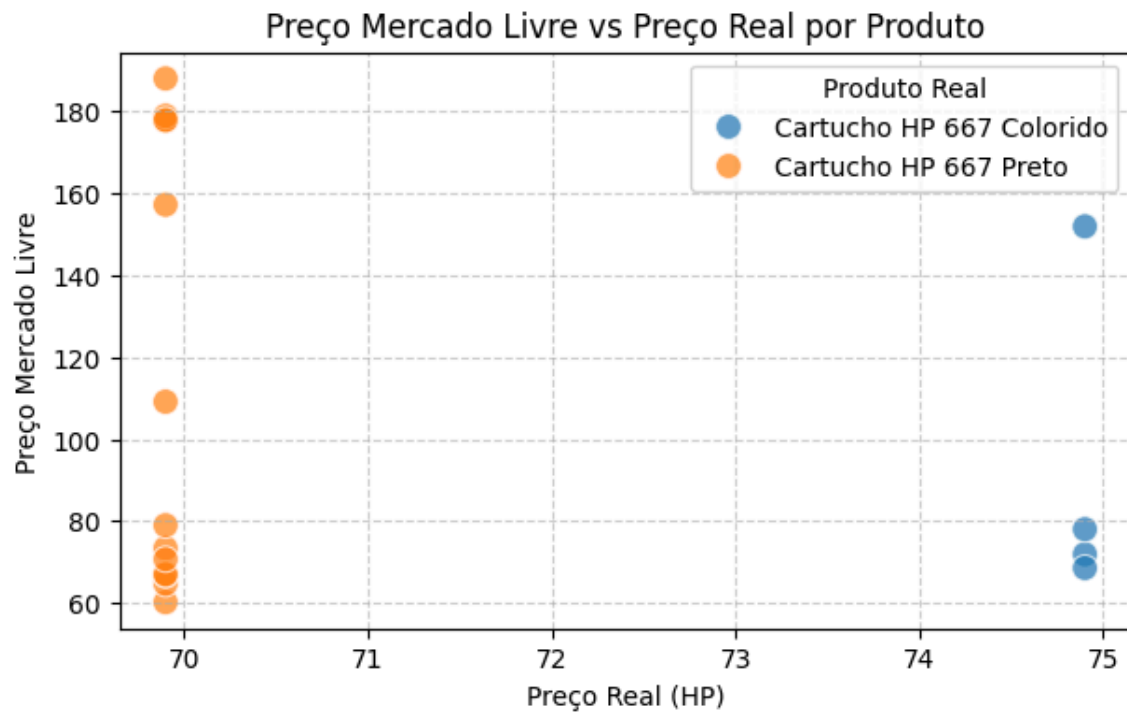
- Utilizamos arquivo CSV enviado pela HP com dados consolidados

## Análise Descritiva e Visual dos Dados

### Distribuição de Preços



## Preço Mercado Livre vs Preço Real por Produto



## Distribuição de Preço do Mercado Livre por Produto Real

