

Eletrôportáteis da Casas Bahia 5 estrelas

Filtragem
por Categoria

Ar-Condicionado

Aspirador

Batedeira

Bebedouro

Blender

Produtos Disponíveis com maiores avaliações

id_produto	avaliacao	qtd_avaliacoes	marca	tipo	preco
10	5.0	14173	Mondial	Aspirador	141,01
1340	4.5	7664	Electrolux	Aspirador	239,90
175	5.0	4069	Mondial	Aspirador	204,90
14	5.0	4006	Electrolux	Aspirador	179,99
1337	4.5	3291	Wap	Aspirador	285,90
1360	4.5	2554	Electrolux	Aspirador	280,00
47	5.0	2503	Electrolux	Aspirador	169,90
1347	4.5	2447	Electrolux	Aspirador	259,00
93	5.0	1999	Electrolux	Aspirador	229,90
40	5.0	1776	Electrolux	Aspirador	215,90

Total de Produtos por Marca

marca	Contagem de produtos
Electrolux	39
Wap	12
Philco	9
Outras	8
Mondial	7
Britânia	5
Oster	4
Tramontina	3
Black	2
Elgin	1
Total	94

Total de vendas por marca



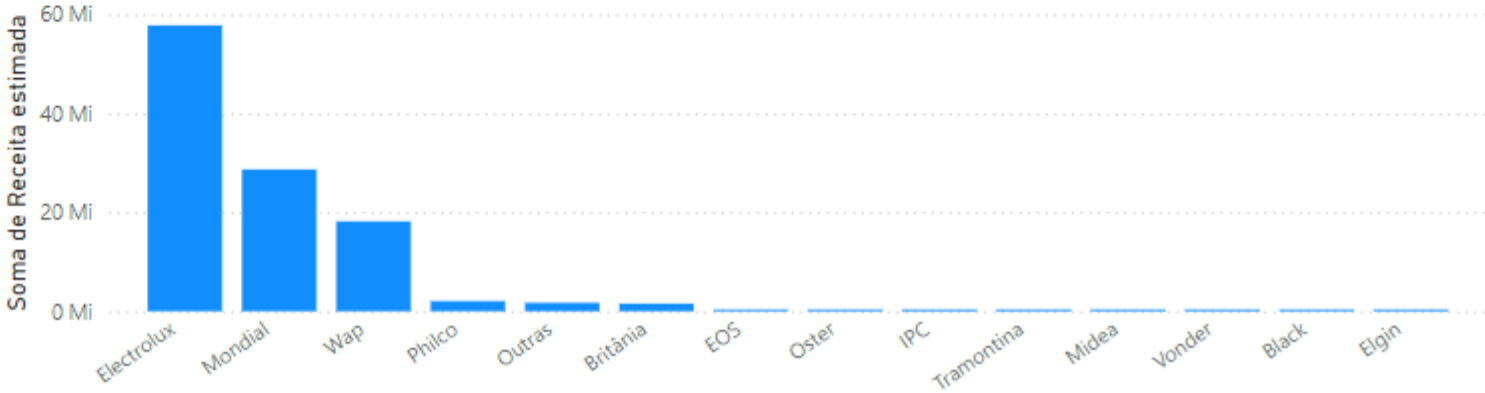
110 Mi

Soma de Receita estimada total

50 Mil

Soma de vendas estimadas total

Soma de Receita estimada por marca



rr6otovf7

October 7, 2024

Análise de Dados de Produtos Eletroportáteis: Um Estudo de Web Scraping com Selenium e Power BI

Autor: César Mâncio Silva

Email: cesarmanciosilva@gmail.com

Telefone: (13) 97403-0583

Este documento apresenta um projeto de web scraping desenvolvido com a biblioteca Selenium em Python, utilizado para coletar dados de produtos de eletroportáteis disponíveis para venda no site da Casas Bahia. Web scraping é uma técnica que permite extrair informações de páginas da web de maneira automatizada, facilitando a análise de dados que são geralmente acessíveis apenas através de navegação em sites. Esse tipo de procedimento é amplamente utilizado em diversas áreas, como marketing, pesquisa de mercado e ciência de dados.

É importante destacar que, neste exemplo, o algoritmo foi ajustado para realizar a extração de dados de apenas 5 páginas de produtos, com o intuito de demonstrar de forma simplificada o processo de coleta. O objetivo deste projeto foi exclusivamente a prática de técnicas de web scraping e análise de dados, e, por essa razão, a coleta total de dados não foi efetuada. No projeto representado no Power BI, foi utilizada uma amostra maior, extraída de 70 páginas de produtos.

Além disso, o processo foi conduzido de maneira ética, respeitando os termos de uso do site e garantindo que a coleta de dados não comprometesse o desempenho do servidor da Casas Bahia ou violasse as regras da plataforma.

```
[ ]: from selenium import webdriver
from webdriver_manager.chrome import ChromeDriverManager
from selenium.webdriver.common.by import By
from selenium.webdriver.chrome.service import Service
from selenium.webdriver.chrome.options import Options
import time
import pandas as pd
servico = Service(ChromeDriverManager().install())

chrome_options = Options()
chrome_options.add_argument("user-agent=Mozilla/5.0 (Windows NT 10.0; Win64;␣
↪x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.121 Safari/537.
↪36")

navegador = webdriver.Chrome(service=servico,options=chrome_options)
```

```

navegador.get('https://www.casasbahia.com.br/c/eletroportateis?
↳&filtro=categoria^d:c73&filtro=avaliacao^d:
↳r5&utm_campaign=DescontoEspecial,DescontoEspecial,DescontoEspecial,DescontoEspecial&utm_med

button = navegador.find_element(By.CLASS_NAME,
↳'styles__SeeMoreButton-sc-d3ac2015-6')
button.click()
brand_section = navegador.find_element(By.CSS_SELECTOR,
↳"[data-testid='brand-filters']")

brands = brand_section.find_elements(By.CLASS_NAME,
↳'styles__LabelText-sc-d3ac2015-4')
marcas = []
for brand in brands:
    marcas.append(brand.text)

print(marcas)

navegador.quit()

cols =
↳['Nome_produto', 'avaliacao', 'qtd_avaliacoes', 'possui_desconto', 'desconto', 'preco', 'marca', '
data = []

for i in range(1,5):
    if i==20:continue
    chrome_options = Options()
    chrome_options.add_argument("user-agent=Mozilla/5.0 (Windows NT 10.0; Win64;
↳ x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/85.0.4183.121 Safari/537.
↳36")

    navegador = webdriver.Chrome(service=servico,options=chrome_options)

    print(f'Pegando dados da pagina {i}...')
    navegador.get(f'https://www.casasbahia.com.br/c/eletroportateis?
↳&filtro=categoria^d:c73&filtro=avaliacao^d:
↳r5&utm_campaign=DescontoEspecial,DescontoEspecial,DescontoEspecial,DescontoEspecial&utm_med
    time.sleep(5)
    cards = navegador.find_elements(By.CLASS_NAME, 'css-gd7wmb')

    for card in cards:
        card_data=[]
        titulo = card.find_element(By.TAG_NAME, 'h3').text
        card_data.append(titulo)

```

```

        avaliacao = float(card.find_element(By.CLASS_NAME, 'css-1vmkvrm').text.
↪split()[-4])
        card_data.append(avaliacao)
        qtd_avaliacoes = float(card.find_element(By.
↪CLASS_NAME, 'product-card__reviews-count-text').text[1:-1])
        card_data.append(qtd_avaliacoes)
        try:
            discount = card.find_element(By.CLASS_NAME, '
↪product-card__discount-wrapper').text.split()

            if len(discount) == 0:
                card_data.append(False)
                discount_prc = "0%"
                card_data.append(discount_prc)
            else:
                card_data.append(True)
                discount_prc = discount[2]
                card_data.append(discount_prc)
        except:
            card_data.append(False)
            discount_prc = "0%"
            card_data.append(discount_prc)
        price = card.find_element(By.
↪CLASS_NAME, 'product-card__highlight-price').text.split()[1].replace(',', '.')
        card_data.append(price)

        confirm_brand=False
        for brand in marcas:
            if brand in titulo or brand.lower() in titulo:
                card_data.append(brand)
                confirm_brand=True
                break
        if confirm_brand==False:
            card_data.append("Outras")
            print(card_data)
        data.append(card_data)
        tipo = titulo.split()[0]
        card_data.append(tipo)
        navegador.quit()
    print(data)
    df = pd.DataFrame(data, columns=cols)
    if 'id_produto' not in df.columns:
        df = df.assign(id_produto=range(1, len(df) + 1))

    # Reordenar as colunas para garantir que 'id_produto' seja a primeira coluna
    df = df[['id_produto'] + [col for col in df.columns if col != 'id_produto']]
    display(df)

```

```
df.to_csv('casas_bahia_eletroportateis.csv', index=False)
```