

Report

March 9, 2023

1 Intellibrand - Data Engineer Test Case

```
[1]: # #####  
# Lucas Lukasavicus Silva  
# March 03, 2023  
# lukasavicus at gmail dot com  
# -----  
# Code Test 4 Data Engineer @ Intellibrand  
# #####
```

Ficamos muito felizes com o seu interesse nesta vaga e esse teste é de grande importância para que possamos avaliar o seu conhecimento técnico, criatividade e organização no desenvolvimento da tarefa. Nosso trabalho diário é desenvolver e implementar soluções de dados que aumentam a performance dos nossos clientes nos principais sites de varejo online. Por isso, criamos este desafio técnico para simular uma atividade que é realizada pelos nossos times especialistas. Nessa tarefa, você deverá encontrar **os 10 produtos com a maior variação de preço, entre o valor encontrado no varejo online (Americanas) e o estipulado pelo fabricante, e os 10 produtos que apresentam maior indisponibilidade**. *Nosso cliente deseja observar com essa task se existem produtos que estão sendo vendidos com valores muito distantes do ideal*. Para isso, disponibilizamos dados coletados nos varejos em diferentes datas para que possa desenvolver essa atividade. Alguns pontos importantes sobre a estrutura dos nossos dados:

- “retailerPrice” – é preço encontrado no varejo;
- “manufacturerPrice” – é preço indicado pelo fabricante;
- “priceVariation” – é a variação do preço, pode ser negativa ou positiva.
- “available” – se o produto está disponível para compra

O escopo do teste deve incluir, mas não se restringindo: - Ingestão dos dados; - Armazenamento dos arquivos; - Tratamento dos dados; - Cálculo das métricas;

Desejáveis: - Aplicação de PyLint ou similar; - Aplicação de Clean Code; - Diagrama da pipeline; - README; - Coverage e Testes unitários;

Os arquivos para o teste estão em formato .zip no link;

<https://drive.google.com/file/d/14nMMC19vpQ8s69TqCkQZh7SxJxlpaOzW/view?usp=sharing>

Esperamos que o resultado seja entregue em um repositório GIT. Boa sorte!

1.1 Import Libraries

```
[2]: import os
from pathlib import Path
import json
import pandas as pd
import numpy as np
from zipfile import ZipFile
import os
```

1.2 Read Data

```
[3]: df = pd.read_parquet('Datatest.parquet.gzip')
```

```
[4]: df.describe()
```

```
[4]:
```

	idRetailerSKU	sk	retailerPrice	retailerRatingCount	\
count	1.009700e+05	100970.000000	22114.000000	30589.000000	
mean	1.122299e+06	50485.939972	591.492001	965.156233	
std	5.857988e+05	29148.099452	2133.545231	4598.041848	
min	2.728400e+04	1.000000	0.990000	0.000000	
25%	6.032950e+05	25243.250000	12.990000	4.000000	
50%	1.008662e+06	50485.500000	25.990000	23.000000	
75%	1.840011e+06	75728.750000	193.662500	200.000000	
max	1.981812e+06	100971.000000	43999.990000	33068.000000	

	retailerFirstImageType	manufacturerExtraImagesTotal	\
count	18023.000000	18096.000000	
mean	0.398657	6.362014	
std	0.489636	5.078741	
min	0.000000	0.000000	
25%	0.000000	3.000000	
50%	0.000000	6.000000	
75%	1.000000	8.000000	
max	1.000000	37.000000	

	manufacturerImagesTotal	retailerExtraImagesTotal	retailerImagesTotal	\
count	18096.000000	18096.000000	18096.000000	
mean	7.362622	4.771883	5.769341	
std	5.076059	6.229312	6.229139	
min	1.000000	0.000000	1.000000	
25%	4.000000	0.000000	1.000000	
50%	7.000000	3.000000	4.000000	
75%	9.000000	7.000000	8.000000	
max	38.000000	47.000000	48.000000	

	descriptionPercentage	manufacturerPrice	priceVariation	\
--	-----------------------	-------------------	----------------	---

count	19066.000000	22114.000000	19979.000000
mean	60.199725	625.989819	1783.011332
std	35.735667	2515.074043	17677.522388
min	0.000000	0.000000	-99.260000
25%	26.670000	3.490000	-9.100000
50%	61.760000	19.290000	3.870000
75%	100.000000	159.000000	27.300000
max	100.000000	49999.000000	284400.000000

	titlePercentage	retailerAverageRating	retailerReviewCount \
count	29622.000000	24457.000000	22072.000000
mean	77.455542	4.497134	4.836988
std	21.471794	0.465938	1.834030
min	0.000000	1.000000	1.000000
25%	60.000000	4.300000	4.000000
50%	83.330000	4.600000	6.000000
75%	100.000000	4.800000	6.000000
max	100.000000	5.000000	6.000000

	correctExtraImagesTotal	extraImagesPercentage	retailerFromPrice \
count	17053.000000	17053.000000	2168.000000
mean	3.102856	42.480888	1319.819161
std	3.806713	43.058324	3852.285873
min	0.000000	0.000000	2.290000
25%	0.000000	0.000000	24.990000
50%	2.000000	28.570000	149.945000
75%	6.000000	91.670000	1641.910000
max	23.000000	100.000000	41999.990000

	priceDiscount
count	2168.000000
mean	72.304405
std	235.027508
min	-935.180000
25%	2.000000
50%	12.990000
75%	58.490000
max	2000.000000

1.3 Business Analysis

1.3.1 1. What are the products with most price variation between retailer and manufacturer?

```
[5]: # 1. Produtos com maior variação de preço entre o Retailer e o Manufacturer:

# Product DataFrame
pdf = df[["idRetailerSKU", "retailerPrice", "manufacturerPrice"]].
    ↳groupby(by="idRetailerSKU").mean().reset_index()
pdf['priceVariation'] = pdf['retailerPrice'] - pdf['manufacturerPrice']
pdf['absPriceVariation'] = abs(pdf['priceVariation'])

pdf.sort_values(by=['absPriceVariation', 'priceVariation'], ascending=False).
    ↳head(10)
```

```
[5]:
```

	idRetailerSKU	retailerPrice	manufacturerPrice	priceVariation	\
	3512	1964334	41931.024483	49999.00	-8067.975517
	1318	866361	7947.609048	14732.72	-6785.110952
	3123	1907007	4866.262857	0.00	4866.262857
	3135	1907784	3729.990000	0.00	3729.990000
	2809	1876930	6341.369310	9999.00	-3657.630690
	2804	1876924	11399.990000	14999.00	-3599.010000
	3138	1917908	5420.679655	8999.00	-3578.320345
	2803	1876923	4975.852069	8499.00	-3523.147931
	3134	1907680	3499.990000	0.00	3499.990000
	901	619475	6920.290000	9999.00	-3078.710000

	absPriceVariation
3512	8067.975517
1318	6785.110952
3123	4866.262857
3135	3729.990000
2809	3657.630690
2804	3599.010000
3138	3578.320345
2803	3523.147931
3134	3499.990000
901	3078.710000

1.3.2 2. What are the most unavailable products?

```
[6]: # 2. Produtos com maior indisponibilidade:

df['availability'] = df['available'].astype(int)
df['occurrence'] = 1
```

```

df['unavailability'] = df['unavailable'].astype(int)

# Unavailability DataFrame
udf = df[['idRetailerSKU', 'availability', 'occurence', 'unavailability']].
    ↳groupby(by='idRetailerSKU').sum().reset_index()
# df[['idRetailerSKU', 'unavailability']].groupby(by='idRetailerSKU').sum().
    ↳reset_index().sort_values(by='unavailability', ascending=False).head(20)

udf['unavailability_pctg'] = udf.apply(lambda row : (row['unavailability'] /
    ↳row['occurence'] * 100) if row['occurence'] != 0 else 0 , axis=1)
udf.sort_values(by=['unavailability_pctg', 'unavailability'], ascending=False).
    ↳head(10)

```

```

[6]:      idRetailerSKU  availability  occurence  unavailability \
7          28005             0          29          29
16         28711             0          29          29
22         28885             0          29          29
30         29149             0          29          29
32         29199             0          29          29
34         29273             0          29          29
51         39939             0          29          29
53         40494             0          29          29
56         40759             0          29          29
57         40794             0          29          29

      unavailability_pctg
7          100.0
16         100.0
22         100.0
30         100.0
32         100.0
34         100.0
51         100.0
53         100.0
56         100.0
57         100.0

```

```

[7]: itemsUnavailable = list(udf[udf['unavailability_pctg'] == 100]['idRetailerSKU'])
f = df['idRetailerSKU'].apply(lambda sku: sku in itemsUnavailable)

```

```

[8]: namedItemsUnavailable = np.unique(df[f]['manufacturerTitle'])
print("All products with 100% unavailability", namedItemsUnavailable)

```

All products with 100% unavailability ['2021 Smart TV LG 43" Full HD 43LM6370
WiFi Bluetooth HDR ThinQAI compativel com Inteligncia Artificial'
'2021 Smart Tv Lg 32" Hd 32Lm627B Wifi Bluetooth Hdr Thinqai Comptivel Com
Inteligncia Artificial']

'2021 Smart Tv Lg 70" 4K Uhd 70Up7750 Wifi Bluetooth Hdr Inteligência Artificial Thing Smart Magic Google Alexa'

'2021 Smart Tv Lg 75" 4K Uhd 75Up8050 Wifi Bluetooth Hdr Inteligência Artificial Thing Smart Magic Google Alexa'

'2022 Smart TV LG 77" 4K OLED77C2 Evo 120Hz Mais Brilho G-Sync FreeSync 4x HDMI 2.1 ThinQAI Google Alexa'

'49" LG Ultra HD 4K TV | 49UJ6525' '49" LG Ultra HD 4K TV | 49UJ6565'

'A-Oxitive Noite, Cuidado peeling, Avène - 30ml'

'A-Oxitive Olhos, Cuidado contorno dos olhos suavizante, Avène - 15ml'

'A-Oxitive Sérum, Sérum protetor antioxidante, Avène - 30ml'

'Absorvente Always Noites Tranquilas Cobertura Suave com 32 unidades'

'Absorvente Interno Intimus Mini - 16 unidades'

'Actine Colors FPS 70 Pele Morena, Darrow - 40g'

'Actine Proteor Solar FPS 60 Cor Universal, Darrow-40'

'Actine Protetor Solar FPS 30 Sem Cor, Darrow- 40g'

'Actine Protetor Solar FPS 60 Sem Cor, Darrow-40g'

'Actine Água Micelar Dermatológica, peles oleosas, Darrow - 100ml'

'Android TV LED 50" TCL P615 4K UHD HDR com Wifi e Bluetooth integrados, 3 HDMI, 2 USB, Controle Remoto com Comando por controle de Voz e Google Assistant'

'Ar Condicionado LG DUAL Inverter VOICE 15.000 Frio 220V - S4-Q15JL31A'

'Ar Condicionado Sem Vento Samsung WindFree Plus 12.000 Btus Quente e Frio (220V)'

'Ar Condicionado Split Hi-Wall Samsung WindFree Inverter 22.000 BTU/h Quente/Frio 220v | AR24TSHCBWKNZ'

'Ar Condicionado Split Inverter Samsung 18000 BTUs WindFree Plus Quente/Frio AR18TSEABWKNZ - 220v'

'Ar Condicionado Split Inverter Samsung 9000 BTUs WindFree Quente/Frio AR09TSHCBWKNZ - 220v'

'Ar-condicionado Sem Vento Samsung WindFree Plus - Quente e Frio - 9.000 btus (220V) 2021'

'BRITANIA - MIXER BRITÂNIA BMX350V 350W'

'BRITANIA - MIXER BRITÂNIA INOX MAXX BMX355P 350W'

'Batedeira Planetária Arno Deluxe 600W com Tigela Inox e 8 velocidades Vermelha SX72 220v'

'Batedeira Planetária Arno Super Chef 750W 5 litros 15 velocidades Vermelha KM02 127v'

'Bisnaguinha Pullman 280g' 'CHÁ LEÃO PRETO NATURAL - 10 SACHÊS'

'Cafeteira Filtro Arno Preferita Inox com 750ml de Capacidade e Jarra de Vidro CFPF 127v'

'Cafeteira Filtro Arno Preferita Inox com 750ml de Capacidade e Jarra de Vidro CFPF 220v'

'Cafeteira Mallory Star Wars - 16 xícaras'

'Caixa Acústica Lg Xboom Rn7- Multi Bluetooth, Super Graves, Entrada De Microfone E Guitarra'

'Calculadora de Mesa MV4121 com Visor 12 Dígitos Elgin'

'Cerveja Amstel Lata 350ml'

'Churrasqueira elétrica Britânia BCQ01D antiaderente 1500W'

'Cicalfate Lábios, Bálsamo reparador labial intensivo, Avène - 10ml'

'Cleanance Gel, Gel de limpeza profunda, Avène - 300g'
 'Compacto Dore Fps 50 Avène - Protetor Solar - Avène, 10g'
 'Cooktop 5 bocas Electrolux a Gás 5 Bocas (KE5GP) - Bivolt'
 'Densiage, condicionador anti envelhecimento capilar, Ducray - 200ml'
 'Densiage, shampoo anti envelhecimento capilar, Ducray - 200ml'
 'Doctar Sensi, Shampoo Anticaspa, Darrow 120ml'
 'FOGÃO BRASTEMP 5 BOCAS DUPLO FORNO COR INOX COM BOTÕES REMOVÍVEIS E EXCLUSIVO ARO PROTETOR 110V'
 'FOGÃO BRASTEMP 5 BOCAS DUPLO FORNO COR INOX COM BOTÕES REMOVÍVEIS E EXCLUSIVO ARO PROTETOR 220V'
 'Ferro a Vapor Elétrico Arno Steamgliss Base Antiaderente e Ponta de Precisão FSGP 127v'
 'Ferro a vapor Arno Smartgliss Inteligente com Desligamento Automático FSC1 127v'
 'Fogão Brastemp 4 Bocas de embutir Preto com mesa de vidro e dupla chama - BYO4XAE- BIVOLT'
 'Forno De Micro-Ondas 30L Revestimento Easyclean, Tecnologia I Wave E 17 Receitas Pré-Programadas'
 'Forno De Micro-Ondas Grill 30L 127V Com Grill De Quartzo E Revestimento Easyclean '
 'Forno De Micro-Ondas Grill 30L Com Grill De Quartzo E Revestimento Easy Clean'
 'Forno De Micro-Ondas Solo 30L 220V Com Revestimento Easyclean E Tecnologia I Wave'
 'Forno Elétrico Arno Airfry 5 em 1 com 20L de Capacidade Preto FOR2 220v'
 'Forno Elétrico Arno Airfry 5 em 1 com 20L de Capacidade Vermelho FOR3 127v'
 'Forno Elétrico Arno Turbo Quartzo 10L de Capacidade e Temporizador FOR1 127v'
 'Forno de Embutir Elétrico Brastemp 84 Litros Cor Inox Espelhado com Convecção e Timer Touch 220V'
 'Fralda Bigfral Confort GD 16 Unidades'
 'Fralda Bigfral Confort Tamanho G 8 Unidades'
 'Fralda Bigfral Plus Normal Tamanho G 16 Unidades'
 'Fralda Bigfral Plus Normal Tamanho G 8 Unidades'
 'Fralda Bigfral Plus Normal Tamanho M 18 Unidades'
 'Fralda Bigfral Plus Normal Tamanho XG 7 Unidades'
 'Fralda Huggies Supreme Care RN - 34 Fraldas'
 'Fralda Pom Pom Derma Protek Jumbo Gd 24Un'
 'Fralda Pom Pom Derma Protek Jumbo M 28Un'
 'Fralda Pom Pom Derma Protek Jumbo P 30Un'
 'Fralda Pom Pom Derma Protek Jumbo Xg 20Un'
 'Fralda Pom Pom Derma Protek Jumbo Xxg 18Un'
 'Fralda Pom Pom Shortinho Jumbo M 36 Unidades'
 'Frigideira Grill Antiaderente Rochedo Inova 28cm'
 'Frigobar Consul 80 litros com Gaveta Multiuso - CRC08CB 110V'
 'Fritadeira Sem Óleo Air Fryer Mondial Family NAF-03I 4L - Preto e Inox'
 'Fritadeira Sem Óleo Air Fryer Mondial Grand Family Inox Red AF-50-RI 5 Litros - Vermelho/Inox'
 'Fritadeira sem Óleo AirFry Arno Moderna com Cuba Antiaderente Preta PFRY 127v'
 'GOOGLE TV LED 55" TCL P635 4K UHD HDR com Wifi dual band e Bluetooth

integrados, Controle Remoto com Comando por controle de Voz, Google Assistant e borda fina'

'Geladeira 371 Litros DFN41 Frost Free Branca Freezer Electrolux'

'Geladeira Brastemp Frost Free 540L Cor Inox com Ice Maker Side Inverse BR080AK 220V'

'Geladeira Brastemp Frost Free Duplex 375L Cor Branca com Espaço Adapt Paineleletrônico BRM45HB 220V'

'Geladeira Brastemp Frost Free Duplex 400L Cor Inox com Paineleletrônico Freeze Control BRM54HK 110V'

'Geladeira Brastemp Frost Free Duplex 400L Cor Inox com Paineleletrônico Freeze Control BRM54HK 220V'

'Geladeira Brastemp Inverse 3 Frost Free 419 litros cor Preta com Freeze Control Pro 110V'

'Geladeira Brastemp Inverse 3 Frost Free 419 litros cor Preta com Freeze Control Pro 220V'

'Geladeira Smart Side By Side 601 Litros Com Moist Balance Crisper E Compressor Linear Inverter 220V'

'Geladeira/Refrigerador Frost Free Electrolux 371 litros (DFN41)'

'Geladeira/Refrigerador Frost Free Electrolux 371 litros (DFN41) 110V'

'Geladeira/Refrigerador Frost Free Electrolux 371 litros (DFN41) 220V'

'Geladeira/Refrigerador Frost Free Electrolux Inox 371L (DFX41) 110V'

'Geladeira/Refrigerador Frost Free Electrolux Inox 371L (DFX41) 220V'

'Geladeira/Refrigerador Top Freezer 431L Branco TF55 Electrolux'

'Google TV LED 65" TCL P725 4K UHD HDR DOLBY VISION - ATMOS, HDR10 Wifi , Bluetooth , 3 HDMI, 2\USB, Comando de voz à distância, Google Assistant e borda ultrafina'

'Google TV LED 75" TCL P725 4K UHD HDR DOLBY VISION - ATMOS, HDR10 , Wifi , Bluetooth , 3 HDMI, 2\USB, Comando de voz à distância, Google Assistant e borda ultrafina'

'Hidratante Desodorante NIVEA Firmador Q10 + Vitamina C Todos os Tipos de Pele 400ml'

'Hidratante Facial e Corporal TriXera Nutrition, Avène- 200ml'

'Ketchup Heinz 2kg' 'Klinse, Shampoo Anticaspa, Darrow - 140ml'

'Lava Louças Electrolux 10 Serviços Branca com Função Higienizar Compras (LL10B) 220V'

'Lava Louças Electrolux 14 Serviços Branca com Função Higienizar Compras (LL14B) 110V'

'Lava Louças Electrolux 14 Serviços Branca com Função Higienizar Compras (LL14B) 220V'

'Lava Louças Electrolux 14 Serviços Inox com Função Higienizar (LL14X) 110V'

'Lava Louças Electrolux 14 Serviços Inox com Função Higienizar (LL14X) 220V'

'Lava e Seca Brastemp 10kg/6kg branca Água Quente com Ciclo Tira Manchas Pro e Ciclo 1h Pronto para Vestir 110V'

'Lava e Seca Brastemp 10kg/6kg branca Água Quente com Ciclo Tira Manchas Pro e Ciclo 1h Pronto para Vestir 220V'

'Lava e Seca Brastemp 10kg/6kg prata Água Quente com Ciclo Tira Manchas Pro e Ciclo 1h Pronto para Vestir 220V'

'Lava-Louças Brastemp 10 Serviços Cor Branca com Ciclo Pesado e Delicado

BLF10BB 110V'

'Lava-Louças Brastemp 10 Serviços Cor Inox com Ciclo Pesado e Delicado BLF10BR 110V'

'Lava-Louças Brastemp 10 Serviços Cor Inox com Ciclo Pesado e Delicado BLF10BR 220V'

'Lavadora Smart Lg Vc5 11Kg Branca Com Inteligência Artificial Aidd '

'Lenços Umedecidos Pampers Aroma de Aloe Vera 192 unidades'

'Lg Blu-Ray Player' 'Lg Xboom Cj44' 'Lhd625.Abrallk'

'Liquidificador Arno Power Max 1000W Preto Com 6 Lâminas e Jarra de 3,1L LN55 127v'

'Liquidificador Arno Power Mix Limpa Fácil Off White com Lâminas Removíveis LQ31 220v'

'Liquidificador Arno Power Mix Limpa Fácil Vinho com Lâminas Removíveis LQ32 127v'

'Liquidificador Arno Power Mix Limpa Fácil Vinho com Lâminas Removíveis LQ32 220v'

'Liquidificador Arno Power Mix Plus Preto com Jarra de 2 litros Cristalina LQ20 220v'

'Liquidificador Arno Power Mix Plus Vermelho com Jarra de 2 litros Cristalina LQ21 127v'

'Macarrão Instantâneo Nissin Sabor Galinha Caipira 85G'

'Maionese Heinz Receitas 405g'

'Micro-Ondas Consul 20L Cor Inox com Função Descongelar Design Moderno e Espelhado CM020BF 110V'

'Micro-Ondas Consul 20L Cor Inox com Função Descongelar Design Moderno e Espelhado CM020BF 220V'

'Micro-ondas Brastemp 20 Litros cor Inox Espelhado com Funções Pré-preparo 110V'

'Micro-ondas Brastemp 32 Litros cor cinza espelhado com Menu Gourmet - BMS46AR- 127V'

'Micro-ondas Brastemp 38 Litros cor inox espelhado com Grill - BMJ38AR 220V'

'Micro-ondas Brastemp De Embutir 40 Litros cor inox com Grill - BMO40AR 220V'

'Mixer Mondial Versatile 220v - M-12'

'Mixer Portátil Arno Turbomix Pro 3 em 1 Inox com 400W e Copo Medidor MX31 127v'

'Mixer Portátil Arno Turbomix Pro 3 em 1 Inox com 400W e Copo Medidor MX31 220v'

'Mixer Portátil Turbomix Arno Duo 200W com Copo Medidor MX21 127v'

'Mixer Portátil Turbomix Arno Duo 200W com Copo Medidor MX21 220v'

'Monitor LG 19,5'' LED HD 20MK400H-B"

'Monitor LG Pro Gamer Ultrawide Full HD 29'' - 29UM69G-B"

'Monitor UltraWide LG 34" 21:9 Full HD IPS HDR10 34WK650-W'

'Mostarda Heinz Sachê 192x5g'

'Máquina de Café Espresso Automática Vermelha G1 Lov Basic Tres 220V'

'Máquina de Lavar 16kg Electrolux Perfect Wash Máquina de Cuidar com Cesto Inox e Jet&Clean (LPE16) 110V'

'Máquina de Lavar 16kg Electrolux Perfect Wash Máquina de Cuidar com Cesto Inox e Jet&Clean (LPE16) 220V'

'Máquina de Lavar 16kg Electrolux Turbo Economia, Silenciosa com Jet&Clean e Filtro Fiapos (LAC16) 110V'

'Máquina de Lavar 16kg Electrolux Turbo Economia, Silenciosa com Jet&Clean e Filtro Fiapos (LAC16) 220V'

'Máquina de Lavar Consul 16Kg com Ciclo Edredom Dispenser Limpa Fácil Dosagem Fácil CWL16AB 110V'

'Máquina de Lavar Consul 16Kg com Ciclo Edredom Dispenser Limpa Fácil Dosagem Fácil CWL16AB 220V'

'NIVEA Creme Facial Antissinais 100g' 'NIVEA Creme Facial Nutritivo 100g'

'NIVEA Demaquilante Bifásico 125ml'

'NIVEA Desodorante Antitranspirante Roll On Dry Comfort 50ml'

'NIVEA Desodorante Antitranspirante Roll On Invisible Black & White Clear 50ml'

'NIVEA Hidratante Labial Amora Shine 4,8 g'

'NIVEA Lenços de Limpeza Demaquilantes 3 em 1 Refrescante 25 un.'

'NIVEA Loção Hidratante Óleos Essenciais Coco & Óleo Monoi 200ml'

'NIVEA MEN Espuma de barbear Original Protect 200ml'

'NIVEA MEN Espuma de barbear Sensitive 200ml'

'NIVEA Men Desodorante Aerosol Dry Fresh 150ml'

'NIVEA Men Desodorante Antitranspirante Roll On Original Protect 50ml'

'NIVEA SUN Protetor Solar Protect & Hidrata FPS50 400ml'

'NIVEA Sabonete em Barra Toque de Baunilha 85g'

'NIVEA Sabonete em barra Antibacteriano 3 em 1 85g'

'NIVEA Tônico Adstringente Facial Controle do Brilho 200ml'

'NIVEA Tônico Hidratante Aqua Rose 200ml'

'Nescafé Dolce Gusto Arno Genio S Plus Multi Bebidas Preta DGS2 127v'

'Nutricerat, shampoo para cabelo seco, muito seco e danificado , Ducray- 200ml'

'Nutriol Hidratante Intensivo Dermatológico, perfume suave, Darrow - 200ml'

'Nutriol Hidratante Intensivo Dermatológico. sem perfume, Darrow - 200ml'

'Panela de Pressão Philco PPP04PI Multifuncional 4L'

'Physiolift Sérum, Sérum preenchedor e reparador, Avène - 30ml'

'ROKU TV LED 43" TCL RS520 FHD Wifi dual band, 3 HDMI, 1 USB, com controle por aplicativo'

'Repelente Off Active Loção 100ml' 'SK6FF.ABRALLK'

'Sabonete Pom Pom Leite e Mel 80g'

'Sabonete Pom Pom Óleo de Amêndoas 80g'

'Sanduicheira Arno Inox Compacta com Placas Antiaderentes SACS 127v'

'Sanduicheira Arno Inox Compacta com Placas Antiaderentes SACS 220v'

'Sanduicheira Grill Arno Dual Inox Abertura 180 graus com Placa Antiaderente GNOX 127v'

'Sanduicheira Grill Arno Dual Inox Abertura 180 graus com Placa Antiaderente GNOX 220v'

'Sensinol, shampoo para couro cabeludo sensível ou sensibilizado, Ducray - 200ml'

'Smart TV 65" LG NanoCell TV | 65SJ8000'

'Smart TV LED 32" LG HD HDR Ativo e WebOS 4.0' 'Smart TV LG 55" 55LH5750'

'Smart Tv Lg 32" Hd Hdr Ativo Inteligência Artificial Thing Ai Virtual Surround Sound'

'Soapelle Sabonete Antisséptico Líquido, Darrow- 140ml'

'TCL QLED TV 50" C725 4K UHD GOOGLE TV, DOLBY VISION, HDR10+, com Wifi dual band e Bluetooth integrados, 3 HDMI, 2 USB, Comando de voz à distância, Google Assistant, DOLBY ATMOS e design sem bordas'

'TV FULL HD 49" LG 49LH5100' 'TV FULL HD LG 43LH5100'

'TV LG Triple XD Engine' 'Toalha Umedecida Bigfral 40 Unidades'

'Toalhas Umedecidas Pom Pom Camomila 50 Unidades'

'Toalhas Umedecidas Pom Pom Camomila Pom Pom Leve 100 Pague 90'

'Trio Anacaps Activ+, suplemento vitaminico mineral, Ducray- 90 Cápsulas'

'Vaporizador de roupas Portátil Arno Steam Power HS81 127v'

'Vaporizador de roupas Portátil Arno Steam Power HS81 220v'

'Ventilador de Teto Arno Ultimate Branco com 3 Pás e Controle Remoto VX10 127v'

'Ventilador de Teto Arno Ultimate Branco com 3 Pás e Controle Remoto VX10 220v'

'Vodka Skyy Infusions Pinapple750ml' 'Água de Colônia Pom Pom 100ml']

1.4 Additional Business Analysis

```
[9]: # Futuras análises:

# a. notListed x unavailable
# b. notListed x priceVariation (eu acho q coisas notListed o pessoal joga o
    ↳ preço pra 0 ou pra 1, então dá uma variação maior);
# c. Será que algum tipo de produto específico apresenta variação maior? Ou
    ↳ indisponibilidade? Tentar fazer uma análise com palavras (título) x essas
    ↳ duas variáveis;
```

1.4.1 3. All products that are not listed are unavailable?

```
[10]: # Análise adicional:
# 1. notListed == Unavailable

print("Unavailable == NotListed? ", sum(df['unavailable'] == df['notListed']))
    ↳ == df.shape[0])
print("notListed is in Unavailable? ", all(df[df['notListed'] ==
    ↳ True]['unavailable']))

# R.: Não. E não acredito que isso seja um erro, mas talvez todos que não estão
    ↳ listado deveriam ser marcados como não disponíveis.
```

```
Unavailable == NotListed? False
notListed is in Unavailable? False
```

1.4.2 4. Exist some sort of correlation between price variation and unavailability?

- The hypothesis is: Products with higher price variation are unavailable, because manufacturer set the price to 0.

```
[11]: # Análise adicional:
# 2. notListed == Unavailable

df["priceVar"] = df["retailerPrice"] - df["manufacturerPrice"]
df[['idRetailerSKU', 'notListed', 'unavailable', 'available', "retailerPrice",
    ↪ "manufacturerPrice", "priceVar"]].sort_values(by=['priceVar'],
    ↪ ascending=False).head(10)
```

```
[11]:      idRetailerSKU  notListed  unavailable  available  retailerPrice  \
53969      1907007      False      False      True      4924.28
52805      1907007      False      False      True      4924.28
17949      1907007      False      False      True      4924.28
15692      1907007      False      False      True      4924.28
24159      1907007      False      False      True      4924.28
69617      1907007      False      False      True      4924.28
25887      1907007      False      False      True      4822.75
96508      1907007      False      False      True      4822.75
8832       1907007      False      False      True      4822.75
48680      1907007      False      False      True      4822.75

      manufacturerPrice  priceVar
53969                0.0    4924.28
52805                0.0    4924.28
17949                0.0    4924.28
15692                0.0    4924.28
24159                0.0    4924.28
69617                0.0    4924.28
25887                0.0    4822.75
96508                0.0    4822.75
8832                 0.0    4822.75
48680                0.0    4822.75
```

```
[12]: shortTitle = df['manufacturerTitle'].apply(lambda title : title.split(' ')[0]
    ↪ if title is not None else '')
```

1.4.3 5. Which are the most unavailable products? (analysis from product title)

```
[13]: productUnavailabilityDf = pd.DataFrame(columns=['shortTitle',
    ↪ 'unavailability'], data=list(zip(shortTitle, df['unavailability'])))
productUnavailabilityDf.groupby('shortTitle').sum().reset_index().
    ↪ sort_values(by=['unavailability'], ascending=False).head(10)
```

```
[13]:      shortTitle  unavailability
124      NIVEA      685
77      Geladeira      498
71      Fralda      405
70      Forno      370
```

78	Geladeira/Refrigerador	275
98	Lava	273
122	Máquina	238
106	Liquidificador	236
1	2021	185
9	Actine	162

1.4.4 6. Which are the products with most price variation? (analysis from product title)

```
[15]: productPriceVarDf = pd.DataFrame(columns=['shortTitle', 'priceVar', 'item'],
      ↪ data=list(zip(shortTitle, df['priceVar'], list(range(df.shape[0])))))

      #productPriceVarDf.groupby('shortTitle').mean().reset_index().
      ↪ sort_values(by=['priceVar'], ascending=False).head(20)

productPriceVarDf.groupby('shortTitle').agg({'priceVar': 'mean', 'item':
      ↪ 'count'}).reset_index().sort_values(by=['priceVar'], ascending=False).
      ↪ head(10)
```

```
[15]:
```

	shortTitle	priceVar	item
85	Home	2357.813529	58
112	Maquina	1844.982000	29
100	Lavadora	1442.297692	57
78	Geladeira/Refrigerador	999.418571	366
116	Microondas	682.400000	56
67	Fogão	586.388571	395
158	Smart	474.002291	301
68	Fone	431.040345	29
5	AR	360.554483	29
122	Máquina	240.990802	799

```
[ ]:
```