

```
In [ ]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
In [ ]: products = pd.read_csv("products.csv")
```

```
In [ ]: products.head()
```

```
Out[ ]:
```

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|---|------------|--------|--|-------------|---|-------------------|-------|
| 0 | 7349 | 165774 | Pañales CHELINO LOVE Talla3 4-10 KG 36unidades | Chelino | <p>Los pañales Chelino, que p... | Infantil | http: |
| 1 | 7347 | 165776 | Pañales CHELINO LOVE Talla5 13-18 KG 30unidades | Chelino | <p>Los pañales Chelino que presentan un novedo... | Infantil | http: |
| 2 | 50282 | 230154 | Pañales Pingo Maxi T4 40 uds | Pingo | <p>Pingo es el primer pañal ecológico con cuat... | Infantil | http: |
| 3 | 7348 | 165775 | Pañales CHELINO LOVE Talla4 9-15 KG 34unidades | Chelino | <p>Los pañales Chelino, que presentan un noved... | Infantil | http: |
| 4 | 24517 | 011905 | Dodot Pañal T4 9-15Kg 30unds | Dodot | <p>¿Qué le pides a un pañal? ¿Absorción máxima... | NaN | http: |

```
In [ ]: products.shape
```

```
Out[ ]: (27953, 7)
```

```
In [ ]: products.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27953 entries, 0 to 27952
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   product_id            27953 non-null  int64
1   sku                   27953 non-null  object
2   name                  27939 non-null  object
3   marca_value           27921 non-null  object
4   short_description     27930 non-null  object
5   analytic_category     24416 non-null  object
6   picture               27947 non-null  object
dtypes: int64(1), object(6)
memory usage: 1.5+ MB
```

```
In [ ]: products[products.duplicated()].shape
```

```
Out[ ]: (0, 7)
```

id y sku

```
In [ ]: products['product_id'].value_counts()[:3]
```

```
Out[ ]: 7349      1
45310      1
45328      1
Name: product_id, dtype: int64
```

```
In [ ]: products[products.duplicated(["product_id"])].head()
```

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|------|------------|--------|---|-------------|---|-------------------|---|
| 3067 | 5572 | 358838 | Aceite Corporal de Calendula Bebe Weleda 200ml | Weleda | <p>Tratamiento protector para pieles sensibles... | Infantil | https://www... |
| 3068 | 4629 | 358820 | Crema Facial Bebe Calendula Weleda 50ml | Weleda | <p>Crema facial para bebés rica en extracto de... | Infantil | https://www... |
| 3097 | 5574 | 150402 | Champu y Gel de Ducha de Calendula Bebe Weleda... | Weleda | <p>Limpieza extra suave para cabello y piel de... | Infantil | https://www... |
| 3098 | 5573 | 357905 | Leche Corporal de Calendula Bebe Weleda 200ml | Weleda | <p>Leche corporal indicada para el cuidado dia... | Infantil | https://www... |
| 3269 | 5575 | 185501 | Baño de Crema de Calendula Bebe 200ml | Weleda | <p>Crema de caléndula ideal para los primeros ... | Infantil | https://www... |

```
In [ ]: products.drop_duplicates(subset="product_id", keep="first", inplace=True)
```

```
In [ ]: products[products.duplicated(["product_id"])].shape
```

```
Out[ ]: (0, 7)
```

```
In [ ]: products.shape
```

```
Out[ ]: (23183, 7)
```

```
In [ ]: products[products["product_id"]== 5572]
```

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|------|------------|--------|--|-------------|---|-------------------|---|
| 1806 | 5572 | 358838 | Weleda Bebé Aceite de Caléndula 200 ml | Weleda | <p>Tratamiento protector para pieles sensibles... | Infantil | https://www... |

```
In [ ]: sum(products["product_id"].value_counts() > 1)
```

```
Out[ ]: 0
```

```
In [ ]: sum(products["product_id"].value_counts() == 1)
```

```
Out[ ]: 23183
```

```
In [ ]: sum(products["sku"].value_counts() == 1)
```

```
Out[ ]: 23173
```

```
In [ ]: sum(products["sku"].value_counts() > 1)
```

```
Out[ ]: 5
```

```
In [ ]: products[products.duplicated(["sku"])]
```

```
Out[ ]:
```

| | product_id | | sku | name | marca_value | short_description | anal |
|--|------------|------|--------------------------|--|-------------|--|------|
| | 7452 | 7194 | sebamed.champu-361303 | Sebamed Champú Dermatológico Anticaspa 500 ml | Sebamed | <p>Champú sin jabón anticasca para el lavado d... | |
| | 11091 | 678 | a3-200181 | Avene Couvrance Compacto Arena 10gr | Avene | <p>Maquillaje compacto en crema para pieles se... | |
| | 11464 | 8337 | Phytocolor-PH969 | Phytocolor Kit Color 6AC Rubio Oscuro Caoba Co... | Phyto | <p>Tinte para el pelo de que no daña el cabell... | |
| | 15308 | 9628 | serv201939.232504-201939 | Sebamed Aceite de Ducha 200 ml | Sebamed | <p>Aceite de ducha, ideal para la higi... | |
| | 25010 | 4525 | 308056.152223-152223 | Femíntim Adulto Gel íntimo Leti | Leti | <p>Gel diario íntimo. \n</p> | |

```
In [ ]: products[products["sku"]== "a3-200181"]
```

```
Out[ ]:
```

| | product_id | | sku | name | marca_value | short_description | analytic_category | |
|--|------------|------|-----------|---|-------------|--|------------------------|-------|
| | 6439 | 4194 | a3-200181 | Avene Couvrance Compacto | Avene | <p>Avene Couvrance Compacto: ... | NaN | https |
| | 11091 | 678 | a3-200181 | Avene Couvrance Compacto Arena 10gr | Avene | <p>Maquillaje compacto en crema para pieles se... | Cosmética y Belleza | https |

OJO. Hay unos SKU repetidos - Qué hacemos?

Ver que los cambios sean iguales en los otros df

```
In [ ]: sum(products["sku"].value_counts() > 1) & sum(products["product_id"].value_c
```

```
Out[ ]: 0
```

```
In [ ]: products[products["product_id"] == 2493]
```

```
Out[ ]:
```

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|------|------------|--------|--|-------------|--|-------------------|----------|
| 2191 | 2493 | 152792 | Crema Reafirmante Woman Isdin 150ml | Isdin | <p>Esta crema está especialmente indicada para... | Infantil | https:// |

pictures

```
In [ ]: products["picture"].isna().sum()
```

```
Out[ ]: 6
```

```
In [ ]: products["picture"].head(6)
```

```
Out[ ]: 0    https://www.mifarma.es/media/catalog/product/1...
1    https://www.mifarma.es/media/catalog/product/1...
2    https://www.mifarma.es/media/catalog/product/2...
3    https://www.mifarma.es/media/catalog/product/1...
4    https://www.mifarma.es/media/catalog/product/0...
5    https://www.mifarma.es/media/catalog/product/0...
Name: picture, dtype: object
```

```
In [ ]: products["picture"].value_counts()[:5]
```

```
Out[ ]: https://www.mifarma.es/media/catalog/productno_selection    3
https://www.mifarma.es/media/catalog/product/1/6/165774_.jpg    1
https://www.mifarma.es/media/catalog/product/2/4/245100_391342_.jpg    1
https://www.mifarma.es/media/catalog/product/1/5/152497_.jpg    1
https://www.mifarma.es/media/catalog/product/8/0/80304935_.jpg    1
Name: picture, dtype: int64
```

```
In [ ]: products[products["picture"] == "https://www.mifarma.es/media/catalog/produ
# Los registros son diferentes. En este caso no sirve esta variable.
```

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|--------------|------------|--------|---|------------------|---|-------------------|------------|
| 15519 | 4464 | 157259 | Pro-Digest 2,6g 20 Capsulas | Nutricion Center | <p>Pro-Digest es un complemento alimenticio na... | Herbolario | https://ww |
| 20402 | 27765 | 154583 | PHB Classic Sensible Cepillo de Dientes | PHB | <p>PHB Classic Sensible\n</p> | Higiene | https://ww |
| 24992 | 22169 | 195156 | Epil-Off 3 Ampollas | Otras Marcas | <p>Reduce y combaten la aparición del vello en... | NaN | https://ww |

In []: `pictures_sep = products["picture"].str.split("/", expand=True)`

In []: `pictures_sep[6].value_counts()[:3]`

Out[]:

```

1    10252
3     1846
2     1376
Name: 6, dtype: int64

```

In []: `pictures_sep[pictures_sep[6].isna()].head()`
#Estos no tienen imag

Out[]:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|--------------|--------|-----|----------------|-------|---------|---------------------|------|------|------|
| 3154 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3720 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 12870 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 15519 | https: | | www.mifarma.es | media | catalog | productno_selection | None | None | None |
| 16328 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |

In []: `pictures_sep[7].value_counts()[:3]`

Out[]:

```

7    3058
6    2943
5    2602
Name: 7, dtype: int64

```

In []: `pic_con_split = products.join(pictures_sep)`

In []: `pic_con_split.head()`

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|---|------------|--------|--|-------------|---|-------------------|-------|
| 0 | 7349 | 165774 | Pañales CHELINO LOVE Talla3 4-10 KG 36unidades | Chelino | <p>Los pañales Chelino, que p... | Infantil | http: |
| 1 | 7347 | 165776 | Pañales CHELINO LOVE Talla5 13-18 KG 30unidades | Chelino | <p>Los pañales Chelino que presentan un novedo... | Infantil | http: |
| 2 | 50282 | 230154 | Pañales Pingo Maxi T4 40 uds | Pingo | <p>Pingo es el primer pañal ecológico con cuat... | Infantil | http: |
| 3 | 7348 | 165775 | Pañales CHELINO LOVE Talla4 9-15 KG 34unidades | Chelino | <p>Los pañales Chelino, que presentan un noved... | Infantil | http: |
| 4 | 24517 | 011905 | Dodot Pañal T4 9-15Kg 30unds | Dodot | <p>¿Qué le pides a un pañal? ¿Absorción máxima... | NaN | http: |

In []: `pic_con_split[pic_con_split[6]=="1"].head()`

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|----|------------|--------|--|--------------------------|---|------------------------|------|
| 0 | 7349 | 165774 | Pañales CHELINO LOVE Talla3 4-10 KG 36unidades | Chelino | <p>Los pañales Chelino, que p... | Infantil | http |
| 1 | 7347 | 165776 | Pañales CHELINO LOVE Talla5 13-18 KG 30unidades | Chelino | <p>Los pañales Chelino que presentan un novedo... | Infantil | http |
| 3 | 7348 | 165775 | Pañales CHELINO LOVE Talla4 9-15 KG 34unidades | Chelino | <p>Los pañales Chelino, que presentan un noved... | Infantil | http |
| 7 | 60743 | 183816 | Cumlaude Sunlaude Niños Ultrafluido Mineral SP... | Rilastil | <p>Muy alta protección solar formulada con <st... | Cosmética y Belleza | http |
| 10 | 75634 | 160842 | FreeLife by Bebe Cash 50 Pañales Talla 4 (7-18kg) | Freelife by bebe cash | <p>Pañales especialmente diseñados para niños ... | Infantil | http |

In []: *#Algo pueden decir los números de las columnas 6 y 7... Pero por lo visto NC*

Short description

In []: `description_sep = products["short_description"].str.split("", expand`In []: `description_sep[description_sep[2].notna()].head()`

| Out[]: | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---------|--|--|---|------|------|------|------|
| 0 | <p>Los pañales | Chelino, que presentan un novedoso di... | 36unidades.\n</p> | None | None | None | None |
| 7 | <p>Muy alta protección solar formulada con | filtros físicos indicada para | bebés desde el primer día y niños/as.... | None | None | None | None |
| 71 | <p> | Mussvital Loción Hidratante es una fó... | Sin parabenos.\n</p> | None | None | None | None |
| 79 | <p>Gel limpiador indicado para | pieles grasas o mixtas y con acné juvenil. </st... | sin jabón garantiza una buena toleran... | None | None | None | None |
| 117 | <p>Blemil Plus 1 Forte es | leche para lactantes indicada para sa... | hasta los 6 meses. 800 gr.\n</p> | None | None | None | None |

```
In [ ]: import re
```

```
In [ ]: products["short_description"] = products["short_description"].str.replace(r'
```

```
In [ ]: products["short_description"].head()
```

```
Out[ ]: 0    Los pañales Chelino, que presentan un novedoso...
1    Los pañales Chelino que presentan un novedoso ...
2    Pingo es el primer pañal ecológico con cuatro ...
3    Los pañales Chelino, que presentan un novedoso...
4    ¿Qué le pides a un pañal? ¿Absorción máxima, q...
Name: short_description, dtype: object
```

names

```
In [ ]: descript_con_split = products.join(description_sep)
```

```
In [ ]: products["name"].isna().sum() #Ya no hay missings en names!
```

```
Out[ ]: 0
```

```
In [ ]: descript_con_split.shape
```

```
Out[ ]: (23183, 14)
```

```
In [ ]: products.shape
```

```
Out[ ]: (23183, 7)
```

```
In [ ]: products[products.duplicated(["name"], keep=False)].head()
```

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|-------------|------------|--------|---|-------------|---|---------------------|---------------------------------|
| 1392 | 15433 | 163575 | Listerine Zero 500ml | Listerine | Enjuague bucal más suave y sin alcohol | Higiene | https:// |
| 1454 | 27786 | 175341 | Singuladerm XpertSun Perfection Natural SPF 50 5g | Singuladerm | Fotoprotector mineral SPF30 en formato polvo, ... | Cosmética y Belleza | https:// |
| 3323 | 49144 | 370406 | Sesderma Hidraloe Crema Contorno Ojos 15 ml | Sesderma | Tu mirada más rejuvenecida e iluminada gracias... | Cosmética y Belleza | https:// |
| 4359 | 4289 | 167926 | Bioderma Atoderm Gel Moussant 500 ml | Bioderma | Gel limpiador diario reengrasante purificante ... | Higiene | https:// |
| 5405 | 47521 | 70407 | Solgar Vitamina C 500 mg 90 comp masticables | Solgar | Vitamina C masticable con sabor naranja. Ayuda... | Nutrición | https:// |

In []: `names_sep = products["name"].str.split(" ", expand=True)`

In []: `names_sep.tail(10)`

Out[]:

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|--------------|------------|------------|----------|------------|----------------|---------|-------------|----------|
| 27931 | Exfoliante | Supernoche | Atashi | 75ml | None | None | None | None |
| 27933 | Soja | Verde | (Mungo) | Bio | Ecosana | 500gr | None | None |
| 27935 | Trío | Sombra | de | Ojos | Hipoalergénica | HYPO | Bell | 09 |
| 27939 | Crema | de | Cara | Pieles | Delicadas | y | Enrojecidas | Bollebio |
| 27941 | Pinzas | de | Depilar | Magnéticas | Punta | Oblicua | U | Tweezers |
| 27942 | Grisi | Pack | Cesta | Aloe | Vera | None | None | None |
| 27943 | Broche | Cadenita | de | Chupete | MAM | Clip | Gris | y |
| 27944 | Espejo | Mifarma | Animal | Print | None | None | None | None |
| 27950 | Corrector | Líquido | Sublime | Ecológico | PuroBio | Tono | 01 | None |
| 27952 | Plantoyo | Cocina | Infantil | de | Madera | None | None | None |

10 rows × 44 columns

In []: `nombre_corto = names_sep[0].astype(str) + " " + names_sep[1].astype(str) + "`

In []: `nombre_corto.tail(10)`

```
Out[ ]: 27931      Exfoliante Supernoche Atashi 75ml None
        27933      Soja Verde (Mungo) Bio Ecosana
        27935      Trío Sombra de Ojos Hipoalergénica
        27939      Crema de Cara Pielles Delicadas
        27941      Pinzas de Depilar Magnéticas Punta
        27942      Grisi Pack Cesta Aloe Vera
        27943      Broche Cadenita de Chupete MAM
        27944      Espejo Mifarma Animal Print None
        27950      Corrector Líquido Sublime Ecológico PuroBio
        27952      Plantoy's Cocina Infantil de Madera
dtype: object
```

```
In [ ]: nombre_corto = nombre_corto.str.replace(" None", "")
```

```
In [ ]: nombre_corto = nombre_corto.str.title()
```

```
In [ ]: nombre_corto.rename("nombre corto");
```

```
In [ ]: nombre_corto = nombre_corto.to_frame().rename(columns= {0: "nombre_corto"})
```

```
In [ ]: #Aquí creamos una nueva columna con los nombres cortos( solo las 5 primeras
products = products.join(nombre_corto)
```

```
In [ ]: products.head()
```

```
Out[ ]:   product_id  sku      name  marca_value  short_description  analytic_category
0      7349  165774  Pañales CHELINO LOVE Talla3 4-10 KG 36unidades  Chelino  Los pañales Chelino, que presentan un novedoso...  Infantil  https://www /mε /
1      7347  165776  Pañales CHELINO LOVE Talla5 13-18 KG 30unidades  Chelino  Los pañales Chelino que presentan un novedoso ...  Infantil  https://www /mε /
2      50282  230154  Pañales Pingo Maxi T4 40 uds  Pingo  Pingo es el primer pañal ecológico con cuatro ...  Infantil  https://www /mε /
3      7348  165775  Pañales CHELINO LOVE Talla4 9-15 KG 34unidades  Chelino  Los pañales Chelino, que presentan un novedoso...  Infantil  https://www /mε /
4      24517  011905  Dodot Pañal T4 9-15Kg 30unds  Dodot  ¿Qué le pides a un pañal? ¿Absorción máxima, q...  NaN  https://www /mε /
```

Ojo! Nombres repetidos... Esto podría ser lo mismo, sin embargo contrastar con los otros df.

Los sku tienen diferencias sutiles. Se podrían eliminar duplicados después de hacer limpieza de sku...

marca

```
In [ ]: products["marca_value"] = products["marca_value"].str.lower()
```

```
In [ ]: products["marca_value"].head()
```

```
Out[ ]: 0    chelino
        1    chelino
        2     pingo
        3    chelino
        4     dodot
        Name: marca_value, dtype: object
```

```
In [ ]: products["marca_value"].isna().sum()
```

```
Out[ ]: 29
```

```
In [ ]: descript_con_split[descript_con_split["marca_value"].isna()].head()
```

| Out[]: | product_id | sku | name | marca_value | short_description | an |
|---------|------------|------|----------------------------|--|-------------------|--|
| | 622 | 9467 | 162269 | Prelox 60 Comprimidos | NaN | Ayuda contra la disfunción erectil.\n |
| | 923 | 7105 | 157899 | Dr Dunner Cinnulin 40 Capsulas | NaN | Cuida tus niveles de azúcar: Complemento alime... |
| | 1029 | 1031 | SERV262253X2-262253-262253 | Interapothek Champu Anticaida 400 ml + 400 ml ... | NaN | Champú formulado con provitamina B5 especialme... |
| | 4376 | 7108 | 152810 | Dr Dunner Menosoy 60 Capsulas | NaN | Complemento alimenticio a base de extracto de ... |
| | 4451 | 513 | 317171x2-317171-317171 | La Roche Posay Solucion Micelar 200ml + 200 ml... | NaN | Solucion Micelar La Roche Posay Desmaquillante... |

```
In [ ]: products["marca_value"][622] = "prelox"
```

/tmp/ipykernel_269695/4007907723.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][622] = "prelox"

```
In [ ]: products["marca_value"][923] = "dr. dunner"  
products["marca_value"][1029] = "interapothek"
```

/tmp/ipykernel_269695/2740116173.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][923] = "dr. dunner"
/tmp/ipykernel_269695/2740116173.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][1029] = "interapothek"

```
In [ ]: products["marca_value"][4376] = "dr. dunner"
products["marca_value"][4451] = "la roche posay"
products["marca_value"][5063] = "medela"
products["marca_value"][5738] = "dr. dunner"
products["marca_value"][5863] = "dr. dunner"
products["marca_value"][5939] = "dr. dunner"
products["marca_value"][7308] = "medela"
products["marca_value"][9188] = "dr. dunner"
products["marca_value"][9820] = "medela"
products["marca_value"][9188] = "dr. dunner"
products["marca_value"][9992] = "medela"
products["marca_value"][10281] = "medela"
products["marca_value"][11338] = "dr. dunner"
products["marca_value"][11680] = "dr. dunner"
products["marca_value"][12641] = "dr. dunner"
products["marca_value"][14110] = "lierac"
products["marca_value"][16104] = "medela"
products["marca_value"][16774] = "mi rebotica"
products["marca_value"][17680] = "mi rebotica"
products["marca_value"][18124] = "dr. dunner"
products["marca_value"][19842] = "medela"
products["marca_value"][20447] = "medela"
products["marca_value"][21037] = "dr. dunner"
products["marca_value"][24622] = "dr. dunner"
products["marca_value"][13094] = "rene furterer"
```

```
/tmp/ipykernel_269695/2746017985.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][4376] = "dr. dunner"
```

```
/tmp/ipykernel_269695/2746017985.py:2: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][4451] = "la roche posay"
```

```
/tmp/ipykernel_269695/2746017985.py:3: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][5063] = "medela"
```

```
/tmp/ipykernel_269695/2746017985.py:4: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][5738] = "dr. dunner"
```

```
/tmp/ipykernel_269695/2746017985.py:5: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][5863] = "dr. dunner"
```

```
/tmp/ipykernel_269695/2746017985.py:6: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][5939] = "dr. dunner"
```

```
/tmp/ipykernel_269695/2746017985.py:7: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][7308] = "medela"
```

```
/tmp/ipykernel_269695/2746017985.py:8: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][9188] = "dr. dunner"
```

```
/tmp/ipykernel_269695/2746017985.py:9: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy  
products["marca_value"][9820] = "medela"
```

```
/tmp/ipykernel_269695/2746017985.py:10: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame
```

```
See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy
```

```
/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][9188] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:11: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][9992] = "medela"
/tmp/ipykernel_269695/2746017985.py:12: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][10281] = "medela"
/tmp/ipykernel_269695/2746017985.py:13: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][11338] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:14: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][11680] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:15: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][12641] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:16: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][14110] = "lierac"
/tmp/ipykernel_269695/2746017985.py:17: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][16104] = "medela"
/tmp/ipykernel_269695/2746017985.py:18: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][16774] = "mi rebotica"
/tmp/ipykernel_269695/2746017985.py:19: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
products["marca_value"][17680] = "mi rebotica"
/tmp/ipykernel_269695/2746017985.py:20: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```


See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][18124] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:21: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][19842] = "medela"
/tmp/ipykernel_269695/2746017985.py:22: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][20447] = "medela"
/tmp/ipykernel_269695/2746017985.py:23: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][21037] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:24: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][24622] = "dr. dunner"
/tmp/ipykernel_269695/2746017985.py:25: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
products["marca_value"][13094] = "rene furterer"

```
In [ ]: products[products["marca_value"] == "rene furterer"].head()
```

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|------|------------|-------------|--|---------------|---|------------------------|------|
| 211 | 8704 | 239442 | RENE furterer triphasic vht+ 8 frascos 5,5ml | rene furterer | Triphasic VHT contrarresta la caída del cabell... | Cosmética y Belleza | http |
| 258 | 8650 | 150710 | RENE furterer okara champu activador luz 200ml | rene furterer | Champú activador de luz indicado para lucir un... | Cosmética y Belleza | http |
| 605 | 8642 | 151105 | RENE furterer naturia champu seco 75ml | rene furterer | Champú en seco que absorbe la grasa y las impu... | Cosmética y Belleza | http |
| 966 | 46294 | 150712duplo | RENE Furterer Initia Champu Volumen Vitalidad ... | rene furterer | Champú de uso frecuente indicado para la higie... | Higiene | http |
| 1006 | 8705 | 239434 | RENE furterer rf 80 tratamiento anticaida 12 a... | rene furterer | Tratamiento anticaída con una doble fórmula pa... | Cosmética y Belleza | http |

In []: `products[products["marca_value"].isna()]`

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|-------|------------|--------|--|-------------|--|------------------------|-------|
| 16276 | 5664 | 160250 | Flammaspray 75ml | NaN | Quemaduras del Sol\n | Cosmética y Belleza | https |
| 26700 | 96969 | 190406 | Loción Capilar Anticaída Forte Farline 100ml | NaN | Loción capilar anticaída concentrada indicada ... | Cosmética y Belleza | https |

Quedan dos nulos en marca. A ver qué se puede hacer con ellos

In []: `products[products["marca_value"] == "interapothek"].head()`

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|-----|------------|--------|---|--------------|---|-------------------|-----------|
| 132 | 10090 | 375642 | ia Bastoncillos de Seguridad 50ud | interapothek | Bastoncillos de algodón hidrófilo para la higi... | Infantil | https://w |
| 198 | 898 | 156864 | Interapothek Solucion Unica 500ml | interapothek | Interapothek Solucion Unica 500ml\n | Óptica | https://w |
| 318 | 956 | 328230 | Interapothek Esponja Natural Grande | interapothek | La esponja natural de Interapothek limpia con ... | Infantil | https://w |
| 580 | 9506 | 165981 | Interapothek Agua de Mar 100 ml | interapothek | Agua de mar\n | Higiene | https://w |
| 784 | 28095 | 482307 | Interapothek Compresas Gasa Esteril c/50 | interapothek | 50 gasas esteriles para limpiar cualquier clas... | Higiene | https://w |

category

In []: `products["analytic_category"].value_counts()`

Out[]:

| | |
|---------------------|------|
| Cosmética y Belleza | 6821 |
| Higiene | 4334 |
| Infantil | 3857 |
| Herbolario | 2487 |
| Nutrición | 1461 |
| Ortopedia | 496 |
| Vida Íntima | 259 |
| Óptica | 195 |
| Perfumeria | 145 |
| Veterinaria | 45 |

Name: analytic_category, dtype: int64

In []: `products["analytic_category"].unique()`

Out[]: `array(['Infantil', nan, 'Cosmética y Belleza', 'Óptica', 'Higiene',
'Herbolario', 'Ortopedia', 'Nutrición', 'Vida Íntima',
'Veterinaria', 'Perfumeria'], dtype=object)`

In []: `products["analytic_category"].isna().sum()`

Out[]: 3083

In []: `# Las categorías se pueden tomar del otro df`

In []: `products[products["analytic_category"].isna()].head()`

Out[]:

| | product_id | sku | name | marca_value | short_description | analytic_category | |
|----|------------|--------|---|-------------|---|-------------------|--|
| 4 | 24517 | 011905 | Dodot Pañal T4 9-15Kg 30unds | dodot | ¿Qué le pides a un pañal? ¿Absorción máxima, q... | NaN | https://www.1 /mei/1 |
| 5 | 11938 | 011896 | Dodot Activity T3 De 4 A 10 Kg 56 Unidades | dodot | Pañales que mantienen la piel del culito seca ... | NaN | https://www.1 /mei/1 |
| 6 | 13109 | 011894 | Dodot Pañal Sensitive T1 2-5Kg 30Unds | dodot | Pañales para recién nacido que evitan la irrit... | NaN | https://www.1 /mei/1 |
| 14 | 12051 | 013687 | Dodot Sensitive Pañal Primeras Semanas T/0 Has... | dodot | Pañales Dodot para una máxima absorción de pip... | NaN | https://www.1 /mei/1 |
| 31 | 13122 | 011907 | Dodot Pañal Pack Semanal T/5 13-18kg 25ud | dodot | Pañales pack semanal\n | NaN | https://www.1 /mei/1 |

product_id - sku

Varios duplicados. El número es diferente al de sku duplicados. product_id conecta con órdenes. Importante verificar que hacen match.

sku

IMPORTANTE. El sku nos va a permitir integrar con dataset de categorías.

category

Hay 3537 nulos... Tomar del otro dataset y corregir. Cubrimos las categorías primarias... Se podría tirar un modelo con palabras para obtener categorías secundarias? Quizás no sea necesario. Pensar...

In []: `products.head()`

| Out[]: | product_id | sku | name | marca_value | short_description | analytic_category | |
|---------|------------|--------|--|-------------|--|-------------------|---|
| 0 | 7349 | 165774 | Pañales CHELINO LOVE Talla3 4-10 KG 36unidades | chelino | Los pañales Chelino, que presentan un novedoso... | Infantil | https://www.../m.../ |
| 1 | 7347 | 165776 | Pañales CHELINO LOVE Talla5 13-18 KG 30unidades | chelino | Los pañales Chelino que presentan un novedoso ... | Infantil | https://www.../m.../ |
| 2 | 50282 | 230154 | Pañales Pingo Maxi T4 40 uds | pingo | Pingo es el primer pañal ecológico con cuatro ... | Infantil | https://www.../m.../ |
| 3 | 7348 | 165775 | Pañales CHELINO LOVE Talla4 9-15 KG 34unidades | chelino | Los pañales Chelino, que presentan un novedoso... | Infantil | https://www.../m.../ |
| 4 | 24517 | 011905 | Dodot Pañal T4 9-15Kg 30unds | dodot | ¿Qué le pides a un pañal? ¿Absorción máxima, q... | NaN | https://www.../m.../ |

```
In [ ]: #Guardo documento:  
products.to_csv("products_limpieza1.csv")
```