

m06_07_regex_parte_IV

April 19, 2021

1 Curso de Python do DS ao DEV

Comunidade DS - Meigarom Lopes

2 Modulo 06 - Python Avançado

3 Introdução a Regex (Expressões Regulares) - Parte IV

```
[43]: import re
import pandas as pd
```

```
[66]: data = pd.read_csv( '/Users/meigarom.lopes/repos/python-ds-ao-dev/Module01/
    ↳products_hm.csv' )

# product id
data = data.dropna( subset=['product_id'] )
data['product_id'] = data['product_id'].astype( int )

# product name
data['product_name'] = data['product_name'].apply( lambda x: x.replace( ' ',
    ↳'_' ).lower() )

# product price
data['product_price'] = data['product_price'].apply( lambda x: x.replace( '$ ',
    ↳'' ) ).astype( float )

# scrapy datetime
data['scrapy_datetime'] = pd.to_datetime( data['scrapy_datetime'],
    ↳format='%Y-%m-%d %H:%M:%S' )

# style id
data['style_id'] = data['style_id'].astype( int )

# color id
data['color_id'] = data['color_id'].astype( int )
```

```

# color name
data['color_name'] = data['color_name'].apply( lambda x: x.replace( ' ', '_' ).
→replace( '/', '_' ).lower() if pd.notnull( x ) else x )

# fit
data['fit'] = data['fit'].apply( lambda x: x.replace( ' ', '_' ).lower() if pd.
→notnull( x ) else x )

# size number
data['size_number'] = data['size'].apply( lambda x: re.search( '\d{3}cm', x ).
→group(0) if pd.notnull( x ) else x )
data['size_number'] = data['size_number'].apply( lambda x: re.search( '\d+', x_
→).group(0) if pd.notnull( x ) else x )

# size model
data['size_model'] = data['size'].str.extract( '(\d+/\d+)' )
data = data.drop( columns=['size', 'product safety'], axis=1 )

# composition
data.head()

```

```

[66]:  product_id product_category product_name  product_price \
0    636207006    men_jeans_slim    slim_jeans             19.99
1    636207006    men_jeans_slim    slim_jeans             19.99
2    636207006    men_jeans_slim    slim_jeans             19.99
3    636207006    men_jeans_slim    slim_jeans             19.99
4    636207006    men_jeans_slim    slim_jeans             19.99

      scrapy_datetime  style_id  color_id  color_name  fit \
0  2021-04-11 17:48:05    636207         6  dark_denim_blue  slim_fit
1  2021-04-11 17:48:05    636207         6  dark_gray_denim  slim_fit
2  2021-04-11 17:48:05    636207         6    denim_blue  slim_fit
3  2021-04-11 17:48:05    636207         6         gray  slim_fit
4  2021-04-11 17:48:05    636207         6  light_denim_blue  slim_fit

      composition  size_number  size_model
0  Cotton 98%, Elastane 2%      NaN      NaN
1  Cotton 98%, Elastane 2%      NaN      NaN
2  Cotton 98%, Elastane 2%      NaN      NaN
3  Cotton 98%, Elastane 2%      NaN      NaN
4  Cotton 98%, Elastane 2%      NaN      NaN

```