

01_prepare_data

September 23, 2019

1 Notebook 1. Processing Pipeline Exploration

1.1 ### By Max Pechyonkin

1.2 Introduction

Note: I understand that you might have limited time. So, if you would rather read only one notebook, rather than all three, please only read the last one, called `02_data_pipeline_refactored.ipynb`, it has the final result. The first two notebooks contained information about how I arrived at that result.

This is an exploratory notebook, where I go step by step and figure out how to deal with this or that feature. While reading this notebook, please also look at the accompanying hand-written digital notes in the PDF format that I submitted with the project. They provide some clarity about why I made certain decisions.

This notebook is not production-ready. The next notebook performs refactoring and data processing in one simple API call: `datacleaner.transform(X)`, which uses all the steps that I devised while working on the data in this notebook.

2 1. Read in CSV, ignore invalid rows

```
[1]: import pandas as pd
import json

DATA_PATH = '7004_1.csv'

df = pd.read_csv(DATA_PATH, error_bad_lines=False, warn_bad_lines=False)
```

```
/Users/fazzl/miniconda3/envs/fastai_dev/lib/python3.7/site-
packages/IPython/core/interactiveshell.py:3058: DtypeWarning: Columns
(20,21,25,29,30,36) have mixed types. Specify dtype option on import or set
low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

```
[2]: df.shape
```

```
[2]: (19315, 48)
```

```
[3]: df.columns
[3]: Index(['id', 'asins', 'brand', 'categories', 'colors', 'count', 'dateAdded',
        'dateUpdated', 'descriptions', 'dimension', 'ean', 'features',
        'flavors', 'imageURLs', 'isbn', 'keys', 'manufacturer',
        'manufacturerNumber', 'merchants', 'name', 'prices.amountMin',
        'prices.amountMax', 'prices.availability', 'prices.color',
        'prices.condition', 'prices.count', 'prices.currency',
        'prices.dateAdded', 'prices.dateSeen', 'prices.flavor', 'prices.isSale',
        'prices.merchant', 'prices.offer', 'prices.returnPolicy',
        'prices.shipping', 'prices.size', 'prices.source', 'prices.sourceURLs',
        'prices.warranty', 'quantities', 'reviews', 'sizes', 'skus',
        'sourceURLs', 'upc', 'vin', 'websiteIDs', 'weight'],
        dtype='object')
```

3 2. Remove rows with all NaNs

```
[4]: all_nan_cols = df.isnull().all()
    all_nan_cols[all_nan_cols]

    all_nan_col_names = all_nan_cols[all_nan_cols].index
    all_nan_col_names

    df = df.drop(all_nan_col_names, axis=1)
```

4 3. Remove rows with NaNs percentage > 99%

```
[5]: def get_nans_percentages(data):
    """
    Get percentages of NaNs for each column.

    Returns Series with index of column names and float values.
    """
    nrows, _ = data.shape
    return (data.isnull().sum(axis=0) / nrows).sort_values(ascending=False)

def get_colnames_with_nans_above_thresh(data, thresh: float):
    """
    Return column names with NaNs proportions above the given level.
    """
    percentages = get_nans_percentages(data)
    result = percentages[percentages > thresh]
    return result.index
```

```
[6]: colnames_with_too_many_nans = get_colnames_with_nans_above_thresh(df, thresh=0.
    →99)
    colnames_with_too_many_nans

    df = df.drop(colnames_with_too_many_nans, axis=1)
```

```
[7]: df.head(10)
```

```
[7]:
```

		id	asins		brand	\
0	AVpfHrJ6ilAPnD_xVXOI	NaN			Josmo	
1	AVpfHrJ6ilAPnD_xVXOI	NaN			Josmo	
2	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
3	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
4	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
5	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
6	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
7	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
8	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			
9	AVpfHsWP1cnluZ0-eVZ7	NaN	SERVUS BY HONEYWELL			

		categories	colors		dateAdded	\
0	Clothing,Shoes,Men's Shoes,All Men's Shoes	NaN			2016-11-07T00:45:12Z	
1	Clothing,Shoes,Men's Shoes,All Men's Shoes	NaN			2016-11-07T00:45:12Z	
2	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
3	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
4	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
5	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
6	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
7	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
8	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	
9	All Men's Shoes,Shoes,Men's Shoes,Clothing	NaN			2016-06-14T04:29:57Z	

		dateUpdated		descriptions	\
0	2016-11-07T00:45:12Z	[{"dateSeen": ["2016-11-07T00:45:12Z"], "sourceU...			
1	2016-11-07T00:45:12Z	[{"dateSeen": ["2016-11-07T00:45:12Z"], "sourceU...			
2	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
3	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
4	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
5	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
6	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
7	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
8	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			
9	2016-07-09T20:26:48Z	[{"dateSeen": ["2016-07-09T20:26:48Z"], "sourceU...			

	dimension	ean	...	prices.returnPolicy	prices.shipping	\
0	NaN	6.993020e+11	...	NaN	NaN	
1	NaN	6.993020e+11	...	NaN	NaN	
2	NaN	NaN	...	NaN	NaN	

3	NaN	NaN	...	NaN	NaN
4	NaN	NaN	...	NaN	NaN
5	NaN	NaN	...	NaN	NaN
6	NaN	NaN	...	NaN	NaN
7	NaN	NaN	...	NaN	NaN
8	NaN	NaN	...	NaN	NaN
9	NaN	NaN	...	NaN	NaN

	prices.size		prices.sourceURLs	reviews	\
0	NaN		https://www.walmart.com/ip/Josmo-8190-Plain-In...	NaN	
1	NaN		https://www.walmart.com/ip/Josmo-8190-Plain-In...	NaN	
2	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
3	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
4	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
5	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
6	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
7	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
8	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	
9	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...	NaN	

	sizes	skus		sourceURLs		upc	\
0	NaN	NaN		https://www.walmart.com/ip/Josmo-8190-Plain-In...	6.993020e+11		
1	NaN	NaN		https://www.walmart.com/ip/Josmo-8190-Plain-In...	6.993020e+11		
2	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
3	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
4	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
5	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
6	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
7	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
8	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	
9	NaN	NaN		http://www.walmart.com/ip/Studs-Shoe-Large-Pr...		NaN	

	weight
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN

[10 rows x 37 columns]

5 4. Remove marked columns

Analysis why columns below need to be removed is in the notes PDF accompanying this submission.

Main reasons are:

- features that are id-like, i.e. we don't want to train for a specific product id, this will overfit
- features that I deemed not very relevant - such as dates, even though I could be possible to turn those dates into categorical features like day of the week, holiday, season etc. For now I decided to not put additional effort in there and focus on building a working MVP.
- features having NLP data, such as reviews, name of the product. I could build an NLP feature extractor based on a neural net that could probably work, but to save time I ignore it for now
- URLs to images and webpages. Some of them are not available, or extracting requires additional work that is not feasible given my time constraints for this project.

```
[8]: for cname in df.columns:  
     print(f"      '{cname}',")
```

```
'id',  
'asins',  
'brand',  
'categories',  
'colors',  
'dateAdded',  
'dateUpdated',  
'descriptions',  
'dimension',  
'ean',  
'features',  
'imageURLs',  
'keys',  
'manufacturer',  
'manufacturerNumber',  
'merchants',  
'name',  
'prices.amountMin',  
'prices.amountMax',  
'prices.color',  
'prices.condition',  
'prices.currency',  
'prices.dateAdded',  
'prices.dateSeen',  
'prices.isSale',  
'prices.merchant',  
'prices.offer',  
'prices.returnPolicy',  
'prices.shipping',
```

```

'prices.size',
'prices.sourceURLs',
'reviews',
'sizes',
'skus',
'sourceURLs',
'upc',
'weight',

```

```

[9]: cols_to_remove = [
    'id',
    'asins',
    'dateAdded',
    'dateUpdated',
    'descriptions',
    'ean',
    'features',
    'imageURLs',
    'keys',
    'manufacturerNumber',
    'merchants',
    'name',
    'prices.amountMin',
    'prices.dateAdded',
    'prices.dateSeen',
    'prices.sourceURLs',
    'skus',
    'sourceURLs',
    'upc',
    'weight',
]

df = df.drop(cols_to_remove, axis=1)

```

```
[10]: df
```

```

[10]:
      brand
0      Josmo
1      Josmo
2  SERVUS BY HONEYWELL
3  SERVUS BY HONEYWELL
4  SERVUS BY HONEYWELL
...
19310  Xray, XRay
19311  Fila
19312  Seek Optics
19313  Seek Optics
19314  Seek Optics

      categories \
0  Clothing,Shoes,Men's Shoes,All Men's Shoes
1  Clothing,Shoes,Men's Shoes,All Men's Shoes
2  All Men's Shoes,Shoes,Men's Shoes,Clothing
3  All Men's Shoes,Shoes,Men's Shoes,Clothing
4  All Men's Shoes,Shoes,Men's Shoes,Clothing
...
19310  Clothing, Shoes & Accessories,Men's Shoes,Boots
19311  Clothing,Shoes,Men's Shoes,All Men's Shoes
19312  Women's Sunglasses,Sunglasses,Bags & Accessori...
19313  Women's Sunglasses,Sunglasses,Bags & Accessori...
19314  Women's Sunglasses,Sunglasses,Bags & Accessori...

```

	colors	dimension	manufacturer	prices.amountMax	prices.color \
0	NaN	NaN	NaN	39.89	NaN
1	NaN	NaN	NaN	51.99	NaN
2	NaN	NaN	NaN	40.02	NaN
3	NaN	NaN	NaN	50.31	NaN
4	NaN	NaN	NaN	46.26	NaN
...
19310	Black	NaN	NaN	29.95	NaN
19311	NaN	NaN	unknown	64.95	NaN
19312	NaN	1 in x 1 in x 1 in	NaN	32.48	NaN
19313	NaN	1 in x 1 in x 1 in	NaN	31.48	NaN
19314	NaN	1 in x 1 in x 1 in	NaN	72	NaN

	prices.condition	prices.currency	prices.isSale \
0	NaN	USD	true
1	new	USD	false
2	new	USD	false
3	new	USD	false
4	new	USD	false
...
19310	New with box	USD	False
19311	new	USD	False
19312	NaN	USD	True
19313	NaN	USD	True
19314	NaN	USD	False

	prices.merchant	prices.offer \
0	NaN	REDUCED USD 12.10
1	UnbeatableSale - Walmart.com	REDUCED USD 12.10
2	SIM Supply Inc - Walmart.com	NaN
3	SIM Supply Inc - Walmart.com	NaN
4	SIM Supply Inc - Walmart.com	NaN
...
19310	NaN	NaN
19311	ULTIMATE SHOE ZONE - Walmart.com	NaN
19312	Seek Optics, Inc	NaN
19313	Seek Optics, Inc	NaN
19314	Seek Optics, Inc	NaN

	prices.returnPolicy	prices.shipping	prices.size	reviews	sizes
0	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
...

19310	NaN	USD 7.95	NaN	NaN	12
19311	NaN	NaN	NaN	NaN	NaN
19312	NaN	NaN	NaN	NaN	NaN
19313	NaN	NaN	NaN	NaN	NaN
19314	NaN	NaN	NaN	NaN	NaN

[19315 rows x 17 columns]

5.1 Dealing with currency and price conversion

```
[11]: df['prices.currency'].value_counts()
```

```
[11]: USD                                18524
      AUD                                338
      CAD                                303
      EUR                                107
      GBP                                 22
      UnbeatableSale.com, Inc. - Walmart.com    3
      PalmBeach Jewelry - Walmart.com          3
      Wholesale Costume and Party Supplies - Walmart.com    2
      SHOEBAcca LTD. - Walmart.com             1
      Shoes.com                                1
      Name: prices.currency, dtype: int64
```

First observation - there appear to be some illegal currencies. Second is that there are not only prices in USD, but also in AUD, CAD, EUR and GBP. I could discard that, but I will convert all prices to USD. The best way to do it would be to retrieve the exchange rates for when the prices were published, but I will simplify and take the current rates. In production, this would definitely require more careful analysis.

Another question I would ask is whether features are good predictors in other markets that use other currencies. It might be the case that for a given local market it is better to have a separate model that learn important relationships in the local market.

5.1.1 Currency, Part 1: Remove rows that are not currencies.

```
[12]: currencies_all = df['prices.currency'].value_counts()
      currencies_all
```

```
[12]: USD                                18524
      AUD                                338
      CAD                                303
      EUR                                107
      GBP                                 22
      UnbeatableSale.com, Inc. - Walmart.com    3
      PalmBeach Jewelry - Walmart.com          3
      Wholesale Costume and Party Supplies - Walmart.com    2
      SHOEBAcca LTD. - Walmart.com             1
      Shoes.com                                1
```


Name: prices.currency, dtype: int64

```
[13]: legal_currencies = currencies_all[currencies_all > 20 ].index  
legal_currencies
```

```
[13]: Index(['USD', 'AUD', 'CAD', 'EUR', 'GBP'], dtype='object')
```

```
[14]: 'SHOEBACCA LTD. - Walmart.com' not in legal_currencies
```

```
[14]: True
```

```
[15]: rows_illegal_currency = df['prices.currency'].apply(lambda x: x not in  
    ↳ legal_currencies)  
rows_illegal_fx = rows_illegal_currency[rows_illegal_currency == True].index  
rows_illegal_fx
```

```
[15]: Int64Index([ 7104,  7105,  7106,  7107,  8587,  8588,  8589,  8590,  8591,  
                8592,  8593,  8594,  8595, 12194, 12195, 13076, 13077, 14754,  
                14755, 14756, 14757],  
              dtype='int64')
```

```
[16]: df.shape
```

```
[16]: (19315, 17)
```

```
[17]: df = df.drop(rows_illegal_fx, axis=0)
```

```
[18]: df.shape
```

```
[18]: (19294, 17)
```

```
[19]: df['prices.currency'].value_counts()
```

```
[19]: USD      18524  
AUD        338  
CAD        303  
EUR        107  
GBP         22
```

Name: prices.currency, dtype: int64

5.1.2 Currency, Part 2: calculate the price based on currency and FX rate

- fx_rates is a mapping from dictionary to a scalar that will be multiply price in currency to make it price in dollars
- I used Google to get the rates

```
[20]: fx_rates = {  
    'USD': 1.00,  
    'AUD': 0.68,  
    'CAD': 0.75,  
    'EUR': 1.10,  
    'GBP': 1.25,  
}
```

```
[21]: df['prices.amountMax'].isnull().sum()
```

```
[21]: 0
```

```
[22]: df['prices.currency'].isnull().sum()
```

```
[22]: 0
```

```
[23]: def is_invalid_price_string(s):
    s = str(s)
    illegal_chars = '-T:Z'
    for c in illegal_chars:
        if c in s:
            return True
    return False

illegal_price_mask = df['prices.amountMax'].apply(is_invalid_price_string)
illegal_price_idxes = illegal_price_mask[illegal_price_mask].index
illegal_price_idxes
```

```
[23]: Int64Index([], dtype='int64')
```

As you can see below, these rows clearly don't belong here. Not only the price is illegal, it looks like these are not even shoes.

```
[24]: df.loc[illegal_price_idxes]
```

```
[24]: Empty DataFrame
```

Columns: [brand, categories, colors, dimension, manufacturer, prices.amountMax, prices.color, prices.condition, prices.currency, prices.isSale, prices.merchant, prices.offer, prices.returnPolicy, prices.shipping, prices.size, reviews, sizes]
Index: []

Let's remove these rows.

```
[25]: df = df.drop(illegal_price_idxes)
```

```
[26]: def convert_to_usd(row):
    price = float(row['prices.amountMax'])
    currency = row['prices.currency']
    if not currency:
        # if currency is NaN assume USD and return price
        return price
    fx_rate = fx_rates[currency]
    return price * fx_rate

df['price'] = df.apply(convert_to_usd, axis=1)
df = df.drop(['prices.amountMax'], axis=1)
df
```

```
[26]:
```

	brand	categories \
0	Josmo	Clothing,Shoes,Men's Shoes,All Men's Shoes
1	Josmo	Clothing,Shoes,Men's Shoes,All Men's Shoes
2	SERVUS BY HONEYWELL	All Men's Shoes,Shoes,Men's Shoes,Clothing

3	SERVUS BY HONEYWELL	All Men's Shoes,Shoes,Men's Shoes,Clothing
4	SERVUS BY HONEYWELL	All Men's Shoes,Shoes,Men's Shoes,Clothing
...
19310	Xray, XRay	Clothing, Shoes & Accessories,Men's Shoes,Boots
19311	Fila	Clothing,Shoes,Men's Shoes,All Men's Shoes
19312	Seek Optics	Women's Sunglasses,Sunglasses,Bags & Accessori...
19313	Seek Optics	Women's Sunglasses,Sunglasses,Bags & Accessori...
19314	Seek Optics	Women's Sunglasses,Sunglasses,Bags & Accessori...

	colors	dimension	manufacturer	prices.color	prices.condition	\
0	NaN	NaN	NaN	NaN	NaN	
1	NaN	NaN	NaN	NaN	new	
2	NaN	NaN	NaN	NaN	new	
3	NaN	NaN	NaN	NaN	new	
4	NaN	NaN	NaN	NaN	new	
...	
19310	Black	NaN	NaN	NaN	New with box	
19311	NaN	NaN	unknown	NaN	new	
19312	NaN	1 in x 1 in x 1 in	NaN	NaN	NaN	
19313	NaN	1 in x 1 in x 1 in	NaN	NaN	NaN	
19314	NaN	1 in x 1 in x 1 in	NaN	NaN	NaN	

	prices.currency	prices.isSale	prices.merchant	\
0	USD	true	NaN	
1	USD	false	UnbeatableSale - Walmart.com	
2	USD	false	SIM Supply Inc - Walmart.com	
3	USD	false	SIM Supply Inc - Walmart.com	
4	USD	false	SIM Supply Inc - Walmart.com	
...	
19310	USD	False	NaN	
19311	USD	False	ULTIMATE SHOE ZONE - Walmart.com	
19312	USD	True	Seek Optics, Inc	
19313	USD	True	Seek Optics, Inc	
19314	USD	False	Seek Optics, Inc	

	prices.offer	prices.returnPolicy	prices.shipping	prices.size	\
0	REDUCED USD 12.10	NaN	NaN	NaN	
1	REDUCED USD 12.10	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
...	
19310	NaN	NaN	USD 7.95	NaN	
19311	NaN	NaN	NaN	NaN	
19312	NaN	NaN	NaN	NaN	
19313	NaN	NaN	NaN	NaN	
19314	NaN	NaN	NaN	NaN	

	reviews	sizes	price
0	NaN	NaN	39.89
1	NaN	NaN	51.99
2	NaN	NaN	40.02
3	NaN	NaN	50.31
4	NaN	NaN	46.26
...
19310	NaN	12	29.95
19311	NaN	NaN	64.95
19312	NaN	NaN	32.48
19313	NaN	NaN	31.48
19314	NaN	NaN	72.00

[19294 rows x 17 columns]

5.2 Cleaning isSale column

```
[27]: def get_unique_nans(colname, data, display=True):
        uniques = data[colname].unique()
        nans = data[colname].isnull().sum()
        print(uniques)
        print(f'Uniques: {len(uniques)}')
        print(f'Nans: {nans}')
        if not display:
            return uniques, nans
```

```
[28]: get_unique_nans('prices.isSale', df)
```

```
['true' 'false' True False]
```

```
Uniques: 4
```

```
Nans: 0
```

```
[29]: def process_is_sale(x):
        if isinstance(x, bool):
            return x
        elif x.capitalize() == 'True':
            return True
        elif x.capitalize() == 'False':
            return False
        else:
            print(x)
            print(type(x))
            raise ValueError("Something went wrong!")

df['prices.isSale'] = df['prices.isSale'].apply(process_is_sale)
```

```
[30]: df['prices.isSale']
```

```
[30]: 0      True
      1     False
      2     False
      3     False
      4     False
      ...
     19310    False
     19311    False
     19312     True
     19313     True
     19314    False
      Name: prices.isSale, Length: 19294, dtype: bool
```

5.3 Cleaning categories column

Since categories are presented as comma separated values, let's see how many unique categories are there by splitting them.

After splitting, I realized that the categories are not very informative and decided to remove the column.

```
[31]: all_categories = []

for string in df.categories.unique():
    string = string.lower()
    if 'shoes' not in string:
        print(string)
    subcategories = string.split(',')
    subcategories = [s.strip() for s in subcategories]
    for s in subcategories:
        if s not in all_categories:
            all_categories.append(s)

print(len(all_categories))
```

978

```
[32]: all_categories
```

```
[32]: ['clothing',
      'shoes',
      "men's shoes",
      "all men's shoes",
      "men's casual shoes",
      'accessories',
      'bags',
      'briefcases',
      "men's accessories",
```

"men's halloween costumes",
'adult halloween costumes',
'shoes & accessories',
'halloween',
'men',
'costumes',
'reenactment',
'theater',
'all costumes',
'party & occasions',
"men's clothing",
'underwear',
'shorts',
'sporting goods',
'boxing',
'martial arts & mma',
"men's sunglasses",
"all men's clothing",
"men's outerwear",
'shoes & jewelry',
'team sports',
'basketball',
'backpacks',
'bags & briefcases',
'athletic',
'shoes & handbags',
"men's boots",
'oxfords & derby shoes',
'wingtip',
'jewellery',
'cufflinks & shirt accessories',
'tie clips',
'jewelry',
'religious',
'tie pins & clips',
'ties & cummerbunds',
'bags & accessories',
'soccer',
'shoes & cleats',
"men's running shoes",
'running shoes',
'running gear',
'fitness & exercise equipment',
'athletic & running shoes',
'luggage & bags',
'juniors messenger bags & backpacks',
'home',

'backpacks & bags',
'travel bags',
"men's bags",
'juniors accessories',
'office supplies',
'school supplies',
'boots',
'see more thorogood mens classics brown leather eh safet...',
'slp',
'finish line athletic shoes',
'nike sneakers',
'workwear & uniforms',
"men's workwear & uniforms",
"men's coveralls",
"men's dress shoes",
"women's sunglasses",
"men's work gloves",
'power tools',
'safety apparel',
"women's workwear & uniforms",
'safety & shop gear',
'hand tools',
'tools',
"women's clothing",
"women's work gloves",
'gloves',
"women's workwear accessories",
"men's workwear accessories",
'office safety & security',
"men's socks",
"men's socks & underwear",
'casual',
'home improvement',
'work safety clothing and equipment',
"men's work & safety shoes",
'boat shoes',
'sunglasses',
'walmart vision centers',
'health',
"men's jeans",
'sunglasses & fashion eyewear',
"women's accessories",
"men's activewear",
'scrubs',
'home health care',
'pants',
'hats',

"men's work shoes & boots",
'fashion jewelry',
'fashion rings',
"men's jewelry",
"men's rings",
'necklaces',
'loafers & slip-ons',
'rings',
'men%27s jewelry',
'anniversary jewelry',
'fashion bracelets & charms',
'bracelets',
'men%27s bracelets',
'men%27s anniversary jewelry',
'cross-training shoes',
'athletic apparel',
"men's sleepwear",
'bath robes & wraps',
'personal care',
'bath',
'bed & bath',
'bath towels & rugs',
'football',
'see more nike air jordan future low cool grey black whi...',
"men's slippers",
'baseball & softball',
"men's",
'jeans',
'novelty',
'hoodies',
'novelty & more',
'#37662 in',
'#11998 in',
'#9810 in',
'classics',
'holiday shops',
'halloween store',
'mens costumes',
'contemporary & designer',
'sunglasses & eyewear accessories',
"all women's clothing",
"women's blazers",
'jackets & vests',
'professional & trade workwear',
"men's scrubs",
'medical uniforms',
'slippers',

"men's footwear",
'footwear',
"men's boat & water shoes",
'loafers',
'see more mens genuine vintage leather messenger man han...',
'sneakers',
'industrial',
'safety',
'protective gear',
'occupational',
'golf',
'golf clothing',
'shoes & accs',
"men's golf clothing & shoes",
'socks',
'swimwear',
"all boys' activewear",
'kids',
"boys' clothing",
'fishing',
'outdoor activities',
'fitness & sports',
'sports & outdoors',
"boys' activewear",
'fishing apparel & footwear',
"kids' clothing",
'fitness & cross-training',
"women's messenger bags & backpacks",
'flat front',
'active',
'active shorts',
'clothes',
'trainers',
'wrist watches',
'watches',
'cycling',
'goggles',
"men's shorts",
"see more men's nike zoom hyperrev sz. 10 630913 300",
'laptop & desktop accessories',
'computers/tablets & networking',
'laptop cases & bags',
'consumer electronics',
'gadgets & other electronics',
'video glasses',
'electronics',
'dress shoes',

"men's watches",
'water shoes',
'vests',
'see more saxx 9699 mens vibe orange striped modern fit ...',
"men's underwear",
"men's sport coats & blazers",
"men's suits & sport coats",
'winter sports',
"women's shoes",
"all women's shoes",
"men's oxfords",
"men's work outerwear",
'mixed items',
'wholesale',
'large & small lots',
'see more akademia zero gravity turf shoe 2010 black size...',
"men's pants",
"men's hats",
"men's sandals",
'fashion eyewear/clear glasses',
"men's work shorts",
'see more lee jeans dungarees carpenter fit mens retro s...',
"men's suspenders",
'sports',
'active hoodies',
'nike football cleats',
"men's cleats",
'nike football',
'football cleats',
'cleats',
'wristwatches',
'jewelry & watches',
'fashion necklaces & pendants',
"men's necklaces",
'see more sunglasses ray-ban light rb3550 019 81 64 matt...',
'unisex clothing',
'unisex accessories',
'dress up accessories',
'party supplies',
'active base layers',
'coats & jackets',
'cameras & photo',
'other cameras & photo',
'cowboy boots',
'gloves & mittens',
'hunting',
'tactical & duty gear',

'tactical gloves',
"women's slippers",
'messenger bags',
'safety glasses & face shields',
"men's protective eyewear",
"women's protective eyewear",
"women's work & safety shoes",
"other men's clothing",
'outlet',
'shoes outlet',
'specialty shops',
'swimwear for the family',
"men's swimwear",
'necklaces & pendants',
'belt buckles',
"men's ties & handkerchiefs",
'ebay motors',
'parts & accessories',
'apparel & merchandise',
'motorcycle',
"men's athletic shoes",
'see more timberland premium waterproof boot for men siz...',
"women's watches",
'running',
'steel-toed boots',
"men's steel-toed boots",
'safety clothing',
"men's overalls",
'safety vests & protective clothing',
'unisex',
'see more adult funny animal comical cow humorous bull c...',
'womens costumes',
'oxfords',
'sweatshirts',
'smart watches & accessories',
'smart watches',
'women',
'replacement sunglass lenses',
'dress',
'gifts',
'for him',
"men's work socks & underwear",
'by recipient',
'tools & home improvement',
'safety & security',
'personal protective equipment',
'hand & arm protection',

'lab',
'safety & work gloves',
'safety work gloves',
'scrub bottoms',
'uniforms',
'work & safety',
'medical',
'business & industrial',
'mro & industrial supply',
'coveralls & suits',
'men's sneakers & athletic shoes',
'fan shop',
'team apparel',
'women's boots',
'surf',
'skate & street',
'sports apparel',
'skullies & beanies',
'activewear',
'hats & caps',
'sandals & flip flops',
'see more revo eyewear sunglasses holsby matte dark tort...',
'smart watch accessories',
'road running',
'driver & moccasin',
'fitness',
'running & yoga',
'fashion sneakers',
'walking',
'father's day gifts',
'winter shoes',
'kids' bags',
'dress/formal',
'women's scarves & wraps',
'sandals',
'boating',
'see more dickies 1993snb relaxed fit indigo wash 100 co...',
'trail running',
'skateboarding',
'boxers',
'sneakers & athletic',
'other men's accessories',
'prescription eyewear frames',
'see more bills khakis 100 cotton island twill pleated p...',
'baseball caps',
'trunks',
'batting gloves',

"see more levi's 501 rigid indigo denim big&tall 48x34 s...",
'see more nike hyperfuse supreme sz 10.5 cool grey atomi...',
'fitness gear',
'exercise & fitness',
'sweats & hoodies',
"men's soccer cleats & shoes",
'soccer cleats & shoes',
"young men's clothing",
"young men's shorts",
'other slip-ons',
'sports fan shop',
'shoes fan shop',
'clothing fan shop',
'ties',
'beanies & knit hats',
'blazers & sport coats',
"men's loafers",
'balaclavas',
'active pants',
'hunting footwear',
'home & garden',
'cordless drills',
"men's workwear",
'health & personal care',
'comfort shoes',
'under \$100',
"young men's swimwear",
"young men's jeans",
'socks & underwear',
'new arrivals',
'outdoor recreation',
'all halloween costumes',
'fashion bracelets',
"men's bracelets",
'camping',
'hiking',
'hiking backpacks',
'rucksacks',
'bags & backpacks',
'link',
'snow pants & bibs',
'stability',
'overalls',
"see more men's adidas originals san remo shoes us 10 uk...",
'see more wrangler 13mgs george strait cowboy cut origin...',
'grocery',
'lab coats',

'big & tall clothing',
 'big & tall outerwear',
 'see more shrek forever after gingerbread man halloween ...',
 'water sports',
 'swimwear & safety',
 '18ps',
 'prada',
 'swim',
 'briefs',
 'see more nike lunarglide 7 mens running shoes 11 white ...',
 'health & wellness',
 'diabetic care',
 'fashion hoodies & sweatshirts',
 'men%27s',
 'fashions for him',
 'valentine%27s day',
 "see more carhartt b299 men's canvas khaki relaxed fit p...",
 "see more dickies men's cargo work pants black 32w x 34l",
 'active sweatshirts',
 'jackets',
 'tennis',
 'racquet sports',
 "young men's outerwear",
 'all action sports',
 'all shoes',
 'action sports outlet',
 'luggage',
 'clothing & accessories',
 'costumes & more',
 'big & tall pants',
 'bowling',
 'see more polo ralph lauren aaric rugged suede plaid moc...',
 'see more bates mens zr-6 black leather nylon lace up bo...',
 'rugged',
 "girls' bags & accessories",
 "boys' bags & accessories",
 "boys' backpacks",
 "boys' backpacks & messenger bags",
 "girls' backpacks",
 "girls' bags & handbags",
 "girls' clothing",
 'sunglass lens replacements',
 "men's suits",
 "boys' shoes",
 'baby & kids shoes',
 "young men's socks",
 "young men's socks & underwear",

'see more hanro 3086 cotton superior boxer brief 2 inch ...',
"young men's activewear",
'team sports outlet',
'baseball cleats',
'baseball',
'outdoor equipment',
'hiking shoes',
'camping & hiking',
'see more forum novelties st patricks day leprechaun ove...',
'craftsman workwear & work boots',
'craftsman tools',
'miscellaneous supplies',
'medical gloves & masks',
'first aid',
'outdoor tools & supplies',
'lawn & garden',
'boots under \$150',
'athletic socks',
'see more nike tech fleece 1mm windrunner hoodie red men...',
'sandals & flip-flops',
'flip-flop',
'see more wrangler jeans original fit cowboy cut 13mwz b...',
'see more 100 authentic official adidas kaiser 5 liga so...',
'fitness & cross training',
"women's handbags & bags",
'backpacks & bookbags',
'mixed items & lots',
'see more adidas crazy 1 kobe basketball shoes c75736 sz...',
'suits',
'motorsport',
'athletic & outdoor shoes',
'trekking & hiking footwear',
'walking shoes',
'watch bands',
'umbrellas',
'money clips',
'golf shoes',
'golf & racquet sports',
'golf cleats & shoes',
"men's golf shoes",
'shoes by sport',
'hiking boots',
'outdoor',
'hiking & trekking',
'skin care',
'beauty',
'masks & peels',

'outdoor sports',
'baskets & bins',
'home & kitchen',
'storage & organization',
"see more reebok bb4500 men's size 9d medium",
'see more riggs workwear wrangler mens jeans 36x36 carpe...',
'lacrosse',
'lacrosse cleats',
"see more men's wrangler riggs workwear contractor indig...",
'baby & toddler shoes',
'#404 in',
'#99 in',
'#4',
'393 in',
'see more jack spade roadway messenger laptop field gray...',
'#10772 in',
'#2382 in',
'#52971 in',
'outdoor clothing',
'cycling shoes',
'briefcases & laptop bags',
'see more mens genuine leather vintage laptop backpack r...',
'earrings',
'work boots',
'dance',
"women's sleepwear",
"women's sleepwear & robes",
'intimates & sleepwear',
'see more gucci sunglasses gg 1110 s 8e270 100 authentic',
"see more nike sb hawthorne camo cargo men's shorts sz 3...",
'see more adult mens jason voorhees friday the 13th hall...',
'unassigned',
"see more nike men's aw77 cuffed sweatpants royal black ...",
'tactical clothing',
'other tactical & duty gear',
"see more dickies men's wrinkle resistant stain release ...",
'automotive',
'motorcycle & atv',
'sneakers & athletic shoes',
'scuff',
'sun hats',
'see more bostonian wendell mens leather oxfords shoes 1...',
'hunting boots',
'slowpitch softball',
'see more aeromax adult fire fighter suit costume in tan...',
'work & safety shoes',
'jackets & leathers',

'vintage',
"men's vintage clothing",
'outerwear coats & jackets',
'see more tumi beacon hill black briefcase attache 68517...',
'cap toe',
'see more sensei master adult mens costume halloween sta...',
'industrial & construction',
'western',
'neutral',
"young men's pants",
'workout & outdoor',
'cuff links',
'see more 4 pair #420 imperfect size xl leather driver w...',
'see more wings & horns 9687 mens chesterfield gray wool...',
'jackets & coats',
'neckties',
"women's casual shoes",
'bags shoes',
"men's bags shoes",
'oxford & derby',
'dancewear',
'suspenders',
'braces',
'adult dancewear',
'dance accessories',
"see more tyr sport men's synergy all over racer swim br...",
'see more asics 6674 mens tan signature golf coaches car...',
'see more wrangler cowboy cut official prorodeo competit...',
'character shop',
"men's character shop",
'youth',
"caterpillar men's heavy insulated parka: amazon.ca: clothing & accessories",
'outerwear',
'work utility & safety',
'shops',
'men%27s necklaces',
'see more outback trading co. kodiak mens hat black 100 ...',
'see more revo holsby green water lens - matte grey crys...',
'see more mens lee premium select relaxed straight leg j...',
'keychains',
"girls' shoes",
'hats & headwear',
'skateboarding & longboarding',
'camera & photo accessories',
'cases',
'bags & covers',
"see more merrell j23485 men's cafe bask sol casual shoe...",

'big & tall socks & underwear',
"big & tall men's socks",
'everything else',
'weird stuff',
'totally bizarre',
'ballet & dance',
'wristwatch bands',
'high-visibility clothing',
"men's work pants",
'high-visibility pants',
"men's bottoms",
'athletic supporters',
'marketplace',
'cross training',
'6000tx',
'carrera',
'see more mens nike tiempo genio leather fg soccer cleat...',
'pants & shorts',
"see more men's genuine vintage leather messenger should...",
'handbags',
'cut resistant gloves',
'patio',
'planters',
'outdoor gardening & lawn care',
'pots',
'planters & container accessories',
'moccasin',
'military',
'basketball shoes',
'hunting accessories',
'hunting bags & packs',
'golf gloves',
'industrial & construction boots',
'fame',
'cyrus',
'spy',
'other golf clothing',
'see more adidas real madrid anthem jacket aa1661 grey y...',
'maui jim',
'wassup',
'boating shoes',
'anniversary sale',
'foot care',
'cushions & orthotics',
"see more under armour 1241967 men's black coldgear ua o...",
"see more rock revival men's jeans rp9840a210 lowry a210...",
'see more carrera ca85s aviator sunglasses light havana ...',

'fins',
'footwear & gloves',
'laptop accessories',
'all computer accessories',
'giftable items',
'luggage & suitcases',
'computers & laptops',
'tvs & electronics',
'hardside luggage',
'laptops',
'all giftable items',
'computer accessories',
'laptop cases & computer bags',
'young men's underwear",
'see more wilsons leather crazy horse leather tablet bag...',
'men's winter & rain boots",
'see more michael kors mens navy solid regular fit blaze...',
'chukkas',
'see more saucony kinvara 6 men us 10 green running shoe...',
'vision care',
'eyeglasses care',
'safety equipment & gear',
'seasonal',
'shoe care & accessories',
'#21379 in',
'see more just keep livin straight leg denim dark rinse ...',
'welding equipment',
'welding helmets & safety',
'see more carhartt b13 dst mens 100 cotton denim washed ...',
'boys' character shop",
'apron toe',
'work safety equipment & gear',
'work gloves',
'caps',
'see more official adidas core 15 3 4 training pants m35...',
'men%27s shoes',
'father%27s day gifts',
'see more costa del mar cb 01 ogmglp corbina blackout gr...',
'sweatpants',
'big & tall',
'women's halloween costumes",
'skateboards & gear',
'extreme sports',
'wheeled sports',
'see more asics mens 33-m shoes t538n 9993 onyx silver o...',
'casual socks',
'see more oversized werewolf midnight howler creature re...'

'big & tall overalls',
'construction overalls',
'construction workwear',
'big & tall workwear & uniforms',
'see more under armour mens 13 highlight mc football cle...',
'see more kenox vintage pu leather mens briefcase laptop...',
'enduro',
'oakley',
'lightweight jackets',
'cotton',
'cold weather gloves',
'bikinis',
'54ps',
'prada sport',
'suits & sport coats',
'sport coats & blazers',
'perry ellis linen suit blazer: amazon.ca: clothing & accessories',
'hiking shoes & boots',
'see more nike air foamposite one metallic gold size 11 ...',
'men's hiking boots',
'family recreation',
'home hardware & vents',
'nuts',
'scarves',
'scarves & bandanas',
'canvas shoes',
'costumes & accessories',
'wigs',
'see more men's leather vintage backpack shoulder bag me...',
'eyeglass frames',
'health & beauty',
'see more nike modern tech woven short omega blue wolf g...',
'see more men's genuine leather handbag messenger should...',
'cell',
'see more exofficio give-n-go boxer - men's hemlock l',
'see more levis jeans 560 comfort loose fit tapered leg ...',
'see more nike victory golf polo 2015 xx-large college n...',
'other sporting goods',
'reading glasses',
'scarves & wraps',
'climbing',
'see more invicta men's 16962 reserve analog display swi...',
'boys',
'chukka',
'toys',
'board games',
'games & puzzles',

'protective outerwear',
'work slip-ons',
'bow ties',
'bow ties & cummerbunds',
'kitchen',
'monk strap',
'see more walls fr flame resistant work carpenter blue j...',
'see more 2pr wrangler pro rodeo competition jeans origi...',
'see more vans half cab mens skateboarding shoes black v...',
"see more men's genuine vintage brown leather messenger ...",
'see more mens puma neolux ferrari golf shoes white ross...',
'active & performance',
'shells',
'performance snow boots',
'coats',
'jackets and coats',
'clothing and apparel',
'down & down alternative',
"men's jackets & outerwear",
'household essentials',
'food storage bags & containers',
'engines & engine parts',
'outdoors',
'replacement engines & parts',
'outdoor power equipment',
'baskets',
'bins & crates',
'bins',
'totes & carts',
'paper & plastic',
'closet storage',
'wrestling',
'bath tissue',
'see more nike hypervenomx finale ic indoor soccer shoe ...',
'moccasin slippers',
'lightweight/minimal',
'compression & base layers',
'shoes & footwear',
'slide',
'tennis & racquet sports',
'the sneaker project',
'adidas',
'see more asics mens gel-nimbus 18 lite show black silve...',
'fishing clothing',
'every other thing',
'sleepwear & robes',
'video games & consoles',

'video gaming merchandise',
 'sport protective eyewear',
 'see more dockers 3005 mens comfort khaki black relaxedf...',
 'see more bag s leather messenger men mens genuine black...',
 'baseball pants',
 'see more mens vintage hollywood formal white jacket adu...',
 'track & active jackets',
 'sports & fitness',
 'see more saucony 4955 mens kinvara 6 blue running shoes...',
 'ankle boots',
 'braces & supports',
 'backpacking boots',
 'medicine cabinet',
 'see more g-star mens hadron denim low rise straight fit...',
 'boxer briefs',
 'vision & hearing impaired',
 'daily living aids',
 'all big & tall suits & sport coats',
 'big & tall sport coats & blazers',
 'big & tall suits & sport coats',
 'windbreakers',
 'tactical bags & packs',
 'jewellery & watches',
 'wristbands',
 'dress & trouser socks',
 'see more handmadecart genuine mens auth real leather me...',
 'see more brooks adrenaline gts 15 mens shoe olympic 10 ...',
 "men's underwear & undershirts",
 'tights',
 'mid-calf boots',
 "men's hunting clothing",
 'hunting clothing',
 'fashion pendants & necklaces',
 '#24344 in',
 '#7640 in',
 '#3114 in',
 'key chains',
 'rings & cases',
 'see more authentic police s1800m 703g mirrored green bl...',
 'sleep & lounge',
 'sleep sets',
 'sportswear',
 "see more nike tiempo genio leather fg men's firm ground...",
 "women's sandals",
 "see more under armour men's ua storm surf mania boardsh...",
 'watches outlet',
 'amazon watch warranty',

'watches: amazon global delivery available',
 'see more tourbon women men canvas messenger bag foldabl...',
 'men's thermal underwear',
 'see more mens justin j-max pull on western work boot ma...',
 'see more keen 4720 mens pittsburgh brown leather work b...',
 'plain toe',
 'shoes & bags',
 '#23549 in',
 'professional uniforms',
 'industrial workwear',
 'cufflinks',
 'masks & eye masks',
 'see more columbia insect blocker cargo pant fossil men'...',
 'slides',
 'see more 20 ties closet organizer necktie rack storage ...',
 'overalls & coveralls',
 'industrial & scientific',
 'id & document holders',
 'business & credit card cases',
 'see more deezone aluminum rfid blocking credit card hol...',
 'all atv accessories',
 'auto & tires',
 'atv',
 'motorcycle and atv',
 'see more -nike flex supreme tr 3 men's running cross tr...',
 'see more authentic revo re 1019 01 gy holsby black grap...',
 'see more rockport storm surge plain toe ox mens rain sh...',
 'flame resistant jackets',
 'flame resistant clothing',
 'see more ecosusi messenger bag women designer vintage l...',
 'see more lee modern series relaxed fit bootcut jeans si...',
 'see more baffin 5883 mens trapper green rubber waterpro...',
 'figures & playsets',
 'see more nike hypervenom phantom ii fg soccer cleats sh...',
 'see more nike dri fit cushioned dynamic arch no show ta...',
 'dragon',
 'see more sons of anarchy men's flag reaper backpack bla...',
 'see more wrangler cowboy cut 13wmz original fit jeans s...',
 'snow & ice sports',
 'snowboarding',
 'see more men's riggs workwear by wrangler ripstop range...',
 'robes',
 'see more callaway ft chev blucher reptile golf shoes ca...',
 'shop his',
 'all men's',
 'top rated',
 'shop by frame shape',

'wrap',
'featured',
'new',
'shop brand',
'gucci',
'rimless',
'shop brands',
'oval',
'eyeglass cases',
'see more lee carpenter jeans dungarees straight loose f...',
'58qs',
'#259 in',
'#8',
'676 in',
'#294 in',
'sun',
'self-tanners & bronzers',
'gifts for grads',
"see more asics h5q4n.5301 gel lyte iii monaco blue men'...",
'world apparel',
'cycling clothing',
"see more men's 1 4 merino wool cushion hiking socks dar...",
'see more nike hyperadapt shield golf jacket 2015 blue f...',
'other',
'see more nike hypervenom phantom ii fg purple volt socc...',
'frogskins lx',
'see more crocs classic realtree clogs unisex khaki mens...',
'keyrings & keychains',
'bike accessories',
'toys & games',
'specialty bike accessories',
'ride on toys & safety',
'travel',
'travel accessories',
'other travel accessories',
"men's travel kits",
'see more timberland pro hyperion hiker xl safety toe me...',
'see more rockport 7102 mens brown suede memory foam loa...',
"see more dickies men's original 874 classic fit work pa...",
'3513m aviator flat metal',
'ray-ban',
'#14541 in',
'#1980 in',
'#468 in',
'trekking & hiking',
'see more nike air force 1 af1 ultra flyknit mid multico...',
'other football clothing & accs',

"see more chicago bears authentic men's team logo watch ...",
 'charms',
 'daypacks',
 'see more nike kobe mamba xi basketball backpack light b...',
 'gardening gloves & clothing',
 'penny loafer',
 'see more scala hats weathered shape outback hat 2 color...',
 'see more wrangler 13mwz cowboy cut original fit mens je...',
 'equestrian sports',
 '#8634 in',
 '#10571 in',
 '#33154 in',
 "see more thorogood men's black composite safety toe oxf...",
 "see more nike cortez basic leather '06 mens 316418-012 ...",
 "see more reebok royal bb4500 hi men's basketball shoes ...",
 'chemical resistant gloves',
 'snow boots',
 'sale',
 "see more women's columbia bugaboot plus ii omni heat bo...",
 'paintball & airsoft',
 'paintball',
 'coveralls',
 'workwear',
 'safety gear & equipment',
 'workwear & apparel',
 'tools & hardware',
 'protective workwear',
 'see more 619360-415 jordan jumpman adjustable snapback ...',
 'bike racks & storage',
 'uniforms & work clothing',
 'coveralls & jumpsuits',
 'exotic apparel',
 'cold weather accessories',
 'see more desert digital camo headwrap biker bandana do ...',
 'luggage & travel gear',
 'see more hickey freeman 0438 mens gray wool blend 2pc t...',
 'fleece',
 'tuxedos',
 'see more nike vapor carbon 2014 elite td mens football ...',
 '#57666 in',
 '#13299 in',
 'backpacks & messenger bags',
 '#2656 in',
 'see more nike black air jordan 7 td retro jumpman footb...',
 'see more native eyewear sunglasses adult sidecar hybrid...',
 'sunglasses & eyewear',
 'safety shoes & boots',

```
"women's athletic shoes",
'see more maui jim sunglasses red sands ht432-10m matte ...',
'motorcycle & combat',
'see more black green combo paisley durag headwrap skull...',
'see more dr. shots costume men's adult one size fits mo...',
'#4851 in',
'#5991 in',
'#17429 in',
'see more american eagle outfitters men's original strai...',
'road running shoes',
'see more walls men s work coveralls pre shrunk cotton k...',
'slip-ons',
'driving shoes',
'heated jackets',
'other camping',
'hiking clothing',
'see more 13mgshd wrangler george strait men's jeans ori...',
'see more mens tpu x performance bowling shoes with inte...',
'see more mens western wrangler retro slim boot cut jean...',
'see more filson laptop bag briefcase - tan - padded com...',
'#1184 in',
'#5481 in',
'#30632 in',
'loafer',
'men's jewellery',
'jewellery sets',
'see more asics gel-lyte iii h5p1l-7105 tan leather sued...',
'see more dickies 874 pants mens original fit classic wo...',
'see more men's levi's 505 straight leg regular fit jean...']
```

```
[33]: df = df.drop(['categories'], axis=1)
```

5.4 Cleaning brand column

This one is ready to be categorized.

```
[34]: get_unique_nans('brand', df)
```

```
['Josmo' 'SERVUS BY HONEYWELL' 'NATIVE' ... 'HEAD' 'Swix' 'Xray, XRay']
Uniques: 1951
Nans: 258
```

5.5 Cleaning prices.merchant

Ready for categorization.

```
[35]: get_unique_nans('prices.merchant', df)
```

[nan 'UnbeatableSale - Walmart.com' 'SIM Supply Inc - Walmart.com'
'ApparelSave - Walmart.com' 'BHFO, Inc. - Walmart.com'
'STAND UP Ranchers - Walmart.com' 'Designer Studio Inc. - Walmart.com'
'Virtual Exchanges Inc - Walmart.com' 'Unique Bargains' 'Amazon.ca'
'Mens Bodega UK' 'MensBodega - Walmart.com' '773adam773adam - Ebay.com'
'Shopping Blitz - Walmart.com' 'Brands R US - Walmart.com'
'Bestpricecenter.com' 'AreaTrend - Walmart.com' 'Liquidators'
'SHOEBACCA LTD. - Walmart.com' 'Street Moda - Walmart.com'
'redseafashion - Ebay.com' 'Bigsales' 'Zoro - Walmart.com'
'winnsports - Ebay.com' 'ProSales' 'New Occidental' 'Kmart'
'Shoenerds - Walmart.com' 'RBX Active - Walmart.com'
'Eyewear & Co. - Walmart.com' 'Supply the Home - Walmart.com'
'ProMax Commerce - Walmart.com' 'UnbeatableSale.com, Inc.' 'Sears.com'
'Shoplet - Walmart.com' 'Upgrab - Walmart.com' 'shoezoo - Walmart.com'
'Fore Liquidation - Walmart.com' 'JewelryAdviser'
'SureGrip Footwear - Walmart.com' 'XiuFen' 'Bidlessnow'
'BeltOutlet - Walmart.com' 'Shoe Webster - Walmart.com'
'Metro Jewelry - Walmart.com' 'Maxton & Company - Walmart.com'
'Tender Robes' 'Shoes N Fashion - Walmart.com' 'trendunited - Ebay.com'
'Shoes.com' 'Tags Weekly - Walmart.com' 'shoe spot - Walmart.com'
'Amazon.com' 'FUSE Optics' 'ArtofDeals - Walmart.com'
'Big Deal Hunter - Walmart.com' 'Viomart - Walmart.com'
'ClayPro Products - Walmart.com' 'SneakerSupply - Walmart.com'
'Overstock.com' 'Shoes & More - Walmart.com'
'Boots Plus More - Walmart.com' 'Botach Inc. - Walmart.com' 'East Majik'
'Sockbin' 'SOLE-UNLIMITED - Walmart.com'
'Prairie Horse Supply - Walmart.com' 'Love - Walmart.com'
'Shoebuy.com - Walmart.com' 'Bic Warehouse - Walmart.com'
'Golfetail Inc - Walmart.com' 'USAetail' 'discountgolfshop - Ebay.com'
'Kylin Express CA' 'shoetashoeta - Ebay.com'
'International Private Label Inc - Walmart.com' "Vangoddy's Marketplace"
'impactingthegame - Ebay.com' 'Vision Group' 'optical store'
'Zayda Co LLC - Walmart.com' 'cleats4feets - Ebay.com'
'graduate08in82 - Ebay.com' 'Toynk Toys LLC - Walmart.com' 'iBraggiotti'
'2006happyface - Ebay.com' 'WatchGrabber' 'Watch Grabber - Walmart.com'
'SoftballFans - Walmart.com' 'ULTIMATE SHOE ZONE - Walmart.com'
'the_mclain_trainthe_mclain_train - Ebay.com'
'Magic Sneaker City - Walmart.com' 'AllSportsWearUSA - Walmart.com'
'Perf-Moto' 'Apparel Save - Walmart.com' 'GotApparel' 'AreaTrend'
'PrincetonWatches.com' 'Vanguard Wholesale' 'www.ralphlauren.com'
'Ruze, Inc. - Walmart.com' 'daugbatudaugbatu - Ebay.com'
'21 Motorsports - Walmart.com' 'kpklosetkpkloset - Ebay.com'
'e4Hats.com, Inc.' 'Woodbury Outfitters - Walmart.com'
'stylemixxuk - Ebay.com' 'Sears' 'Story Book Kids' 'DiligentBuy'
'Tasharina Corp - Walmart.com'
'National Brands Liquidation - Walmart.com' 'Taylor Gifts - Walmart.com'
'oh!Dreamy' "Marsino's Costumes and Novelties, LLC. - Walmart.com"
'Pure Costumes - Walmart.com' 'nikki2281nikki2281 - Ebay.com'

'In Season Jewelry' "Ron's Home and Hardware - Walmart.com"
 'suntimes365 - Ebay.com' 'Leather Care Supply - Walmart.com'
 'Luxury Divas - Walmart.com' 'Off Price Direct - Walmart.com'
 'Neobits.com' 'ShopLadder - Walmart.com' 'Factory Advantage LLC'
 'sexygirl_07 - Ebay.com' 'sergo7sergo - Ebay.com'
 'Orthotic Shop - Walmart.com' 'OrthoticShop.com - Walmart.com'
 'bcofo - Ebay.com' 'BlockBuster Costumes, LLC - Walmart.com'
 'Best Service Stores - Walmart.com' 'Hobby Warehouse - Walmart.com'
 'gear-house - Ebay.com' 'DrHotDeal' 'ATAFA - Walmart.com'
 'Step Up - Walmart.com' 'Costumes4less.com - Walmart.com'
 'RIGHT BOOT FORWARD - Walmart.com' 'lacysclosetlacyscloset - Ebay.com'
 'Best Global Trader - Walmart.com' 'Trend Nation LLC - Walmart.com'
 'Blanco Bedding' 'Calendars LLC - Walmart.com' 'Tennisexpress'
 'shoesnstuff15 - Ebay.com' 'Shivam' 'RedsGear - Walmart.com'
 'antonline - Walmart.com' 'Rover Store, Inc. - Walmart.com'
 'Retailers Online' 'Brands Online' 'X Shop - Walmart.com'
 'Dapper World Inc.' '101retrokicks, inc - Walmart.com'
 'Brandshopper, LP - Walmart.com' 'whoa2go5whoa2go5 - Ebay.com'
 'The Roost - Walmart.com' 'SocksMax.com' 'RedsGear.com' 'SGIFashion'
 'Walk Into Fashion - Walmart.com' 'My First Aid Company'
 '7th Avenue Costumes - Walmart.com' 'Future Memories Inc. - Walmart.com'
 'HalloweenCostumes - Walmart.com' 'reccos_3 - Ebay.com'
 'cherrycomeoncherrycomeon - Ebay.com' 'RCTECH PTY(Ltd)' 'Sunshade King'
 'IT Glasses' 'Designer Iwear'
 'Nobles Health Care Product Solutions - Walmart.com' 'DigitUniform'
 'Daily Diamond Deal' 'ml4241 - Ebay.com' 'etailsuperstore'
 'sole_eyewear01 - Ebay.com' 'berkshiresoutlet - Ebay.com' 'Pinelawns'
 'toyworld - Ebay.com' 'allforless2013allforless2013 - Ebay.com'
 'PlumStruck - Walmart.com' 'myteamplanet - Walmart.com' 'Amazon'
 'RVCA Canada' 'Beach City Bike - Walmart.com'
 'spenserkaboomspenserkaboom - Ebay.com' 'Myquickmart - Walmart.com'
 'Easy Fit Brands' 'Sunglass Hut' 'OpticsPlanet - Walmart.com'
 'Ami Ventures Inc' 'FabulouStore' 'Outdoor Sports Stores - Walmart.com'
 'LSD Direct USA - Walmart.com' 'Stuhrling Original'
 'Amazing Sneakers - Walmart.com' 'Peltz Shoes - Walmart.com'
 'Captain Dave Inc - Walmart.com' 'Crazy For Bargains LLC - Walmart.com'
 'Shop The Docks Florida - Walmart.com' 'Drill Spot LLC - Walmart.com'
 'Fast Shipping! - Walmart.com' 'Premier Sports'
 'Wholesale Costume and Party Supplies - Walmart.com'
 'Kryptonite Kollektibles' 'BESTON SHOES - Walmart.com' 'OpticsPlanet.com'
 'FrenchChimp' 'ErgodE' 'SUMMITFASHIONS - Walmart.com'
 'Webzom - Walmart.com' 'Elevatit LLC - Walmart.com' 'Seek Optics, Inc'
 'Jomashop' 'salefolks' 'My Times' 'Fanzz - Walmart.com'
 'Stile - Walmart.com' 'shopfarmandcity.com' 'Nilima Online Stores'
 '2015hotsale2015 - Ebay.com' 'Spiffy Custom Gifts'
 'OneStopFanShop - Walmart.com' 'J'colour Socks"
 'EB Sports Store - Walmart.com' 'Campmor - Walmart.com'
 'theshoes_n_stuffguytheshoes_n_stuffguy - Ebay.com'

'Sunglasses Etc - Walmart.com' 'New York Deals - Walmart.com'
 'CosmeticMall.com - Walmart.com' 'briansh11aubriansh11au - Ebay.com'
 'Fashion House' 'abprulesabprules - Ebay.com'
 'Great Golf Sales - Walmart.com' '*scoobies* - Ebay.com' 'Wayfair.com'
 'Shoe Station Inc. - Walmart.com' 'jrsarenajrsarena - Ebay.com'
 'gj2016gj2016 - Ebay.com' 'lowestfirstsports - Ebay.com'
 'Universo Extremo Boards' 'thesavinglanethesavinglane - Ebay.com'
 'British Orchard Home Fashions'
 'lynnstinkerbelllynnstinkerbell - Ebay.com' 'Ibex Optics'
 'We The People - Walmart.com' 'Life and Home - Walmart.com'
 '@FLOW_RUNNING' 'ohDreamy' 'Blu Moon - Walmart.com' 'Amazon.co.uk'
 'GotApparel - Walmart.com' 'Direct Home Supplies - Walmart.com'
 'Costumes4less - Walmart.com' 'Altatac - Walmart.com' 'passthewatch'
 'ShoeMars - Walmart.com'
 'Personal Touch Health Care Apparel - Walmart.com'
 'Accessorymart - Walmart.com'
 'ji-internationalji-international - Ebay.com' 'undefined'
 'absell08absell08 - Ebay.com' 'brooklynajaysbrooklynajays - Ebay.com'
 'mateofmilk - Ebay.com' 'Movaz Trade' 'lenvasilenvasi - Ebay.com'
 'Sunglasses Shop' 'circlesquaredcacirclesquaredca - Ebay.com' 'netailor1'
 'Top One International Corp.' 'Shipppe - Walmart.com'
 'WorldofWatches - Walmart.com' 'Premium Brand Outlet - Walmart.com'
 'Snow Sport Deals - Walmart.com' 'TVTimeDirect - Walmart.com' 'eshades'
 'Clever Training - Walmart.com' 'whozitswhatzits13 - Ebay.com'
 'Boardparadise' 'prncsamanda - Ebay.com' 'Premier Sports - Walmart.com'
 'Watch Express 123 - Walmart.com' 'Perf-Moto - Walmart.com'
 'Sunglass Boutique' 'S.E.G' 'californiakicks - Ebay.com'
 'Wayfair - Walmart.com' 'AllianceSupply.com' 'Shallom'
 'Walmart store - Walmart.com' 'ISelections' 'Buffer Zone - Walmart.com'
 'Craigs Closet - Walmart.com' 'Carr Surplus LLC - Walmart.com'
 'Firefly Buys - Walmart.com' 'Rod's Western Palace - Walmart.com'
 'Style Like Mine - Walmart.com' 'DiscountShop' 'Style Unlimited'
 'ClicktimeEU' 'Golfio Inc. - Walmart.com' 'SeaofDiamonds.com'
 'SpectrumSuperStore' 'DLY Supplies' 'retailliquidators94 - Ebay.com'
 'La Dolce Vitae, Inc. - Walmart.com' 'SIM Supply'
 'LeatherUp - Walmart.com' 'ballerbandit11 - Ebay.com'
 'Gravity Trading Inc - Walmart.com'
 'UnbeatableSale.com, Inc. - Walmart.com' 'Shopping Zulla' 'Time Savings'
 'E-Dials' 'homedepot.com' 'jacksstuffformenjacksstuffformen - Ebay.com'
 'beeleemoon2288 - Ebay.com' 'Flizzards' 'insteaddo - Ebay.com'
 'GuruFrame' 'kewlshoeskewlshoes - Ebay.com' 'Campus Colors - Walmart.com'
 'DailyShoes - Walmart.com' 'Factory Shoe Shop Olney' 'Shoesbyclaire'
 'KimmyShop - Walmart.com' 'Extrememoto' 'Krisar Clothing'
 'Shop Eddies - Walmart.com' 'Watch Related'
 'ShoesandFashions - Walmart.com' 'bcoutdoors2010 - Ebay.com'
 'H & H Outdoor Outfitter & Sport Shop, Inc - Walmart.com'
 'Edealszone LLC' 'zippdogzippdog - Ebay.com' 'WD Quality - Walmart.com'
 'P26 - Walmart.com' 'Solito Fashion - Walmart.com'

'Old Glory - Walmart.com' 'eGilberts' 'Sneaker Kingdom - Walmart.com'
 'Sneaker Kingdom LLC - Walmart.com' 'shaqalz715shaqalz715 - Ebay.com'
 'topbranditemstopbranditems - Ebay.com'
 'deals2treasuredeals2treasure - Ebay.com'
 'happyfeetnowhappyfeetnow - Ebay.com' 'shoemaniacschoemaniac - Ebay.com'
 'WOLFBIKE' 'ebuy*for*you - Ebay.com' 'A&E DEALS'
 'wholesalepricesnycwholesalepricesnyc - Ebay.com' 'Happy Sun USA'
 'Seven Times Six - Walmart.com'
 'shopdesignerdiscountsshopdesignerdiscounts - Ebay.com'
 'mmasharp - Ebay.com' 'Zoro' 'Mobius Commerce - Walmart.com' 'FastMedia'
 'ACCESSORIES - Walmart.com' 'pro_kicks - Ebay.com'
 'RedTag Sports - Walmart.com' 'happyshop66866 - Ebay.com'
 'motvmtmotvmt - Ebay.com' 'lifeandhome.com - Life and Home'
 'Andrew Stevens Collection INC - Walmart.com'
 'Amazing Attire - Walmart.com'
 'discountclothing2015discountclothing2015 - Ebay.com'
 'Outdoor Equipped - Walmart.com' 'MBH-Shop' 'BestlightCA' 'ApparelBarn'
 'Costume Kingdom - Walmart.com' 'Greater Golf Express - Walmart.com'
 'Sportco Warehouse Sporting Goods - Walmart.com'
 'Universal Textiles & Footwear' 'Ice.com' 'costumesworldwide - Ebay.com'
 'Alliance Supply - Walmart.com' 'SavvyWatch.com'
 'Shoe Wonder - Walmart.com' 'Urbanity Boutique - Walmart.com'
 '3dRose LLC' 'Online Stores LLC - Walmart.com' 'nawtyfox - Ebay.com'
 'Totally Costumes - Walmart.com' 'Nawty Fox - Walmart.com'
 'Authentic Glasses - Walmart.com' 'Watchsavings - Walmart.com'
 'fuscarm vision' 'AMI Ventures Inc - Walmart.com' 'myshoespot - Ebay.com'
 'Mezza' 'One Super Bargain'
 'hole-in-one_boutiquehole-in-one_boutique - Ebay.com'
 'super-cleats - Ebay.com' 'Evelots - Walmart.com'
 'ShoeFabs - Walmart.com' 'Sears Authorized Supplier'
 'Tackxas Saddle & Leather Superstore - Walmart.com'
 'NEONBLOND Jewelry & Accessories' 'Ruze Shoes'
 'hyndmanitehyndmanite - Ebay.com' 'Eyewear-Emporium'
 'Vanguard Wholesale - Walmart.com' 'anotherpairanotherpair - Ebay.com'
 'onyafeet - Ebay.com' 'ShoeOcean' 'Final Score Athletics - Walmart.com'
 'Diplomat-jp' 'Wrapables - Walmart.com' 'Designerbrandsforless'
 'Dreamway Trading, LLC' 'The Athlete's Edge - Walmart.com'
 'SPECIALPLACE - Walmart.com' 'karenqianyun - Ebay.com'
 'wordsunlimited - Ebay.com' 'Clorts Outdoor Shoes'
 'Happy Feet and Comfy Feet Slippers - Walmart.com' 'Jewelryweb.com'
 'Summit Sports - Walmart.com'
 'forsalewholesale2007forsalewholesale2007 - Ebay.com'
 'Direct Premium Buys - Walmart.com' 'usmates2015 - Ebay.com'
 'Unique's Shop' 'FashionStar' 'alina-mao - Ebay.com'
 'The Primrose Lane LLC - Walmart.com' 'Sunbelt Products - Walmart.com'
 'IPTravelers' 'RAJO Watches' 'VMInnovations.com' 'Forza Sports'
 'hecyeahhecyeah - Ebay.com' 'Beach Audio Inc - Walmart.com'
 'causer1causer1 - Ebay.com' 'Bijou.inc' 'OpticsFast Inc'

'lasvegasstore10lasvegasstore10 - Ebay.com'
 'Jarvis Tech Inc. - Walmart.com' 'FOOTWEAR ETC. - Walmart.com' 'Polo&Co'
 'Popular Frames - Walmart.com' 'Hobby Hunters Inc. - Walmart.com'
 'Pwatch' 'Omni Cheer - Walmart.com' 'MRITEMS' 'Great Timing Inc.'
 'Targettime' 'Timepiece Trading CA' 'shengli2015 - Ebay.com'
 'Shoobiz4u - Walmart.com' 'pellor.best.dealspellor.best.deals - Ebay.com'
 'Drill Spot' 'All Volleyball, Inc. - Walmart.com' 'Fantasyard.com'
 'troyers - Ebay.com' 'Buyers Direct - Walmart.com'
 'koreamarketkoreamarket - Ebay.com' 'USA Dawgs, Inc - Walmart.com'
 'premiumgoods - Walmart.com' 'Fashion Square - Walmart.com' 'Fuse Lenses'
 'shadesworldshadesworld - Ebay.com' 'ToughKicks, LLC - Walmart.com'
 'Irvs Luggage' 'sienna belle' 'Shop Inc'
 'realtextilerealtextile - Ebay.com' 'framesdirect.com'
 'luopx6811 - Ebay.com' 'Farm & City Supply - Walmart.com' 'Yu Qian'
 'greenturtle2014 - Ebay.com' 'Walk Into Fashion' 'UrbanInspirations'
 'VirVentures - Walmart.com' 'Kryptonite Collectibles - Walmart.com'
 'Wayfair.com - Walmart.com' 'BulkBuyWholesale'
 'Back40 Trading Company - Walmart.com' 'webportdeals - Ebay.com'
 'Worldofwatches' 'eCopUSA - Walmart.com' 'Mobesano'
 'dusktodown - Ebay.com' 'solewearplus - Ebay.com' 'Keep Up Life'
 'Ablackhorse.com, Inc' 'Pureglare' 'Mwave.com'
 'crazymarksgear - Ebay.com' 'Universal Direct Brands'
 'EyeCareUniverse - Walmart.com' 'watch-shop' 'BrightWatches'
 'Bling Jewelry - Walmart.com' 'Bling Jewelry'
 'Paul's Shoes - Walmart.com' 'Hold-Up Suspender Company - Walmart.com'
 'thoro-bred - Ebay.com' 'thetargetbuys' 'Daulia' 'BETAWATCHES'
 'bonesy - Ebay.com' 'Twinix Store' 'Seen on Tv Inc'
 'sportsauctionssportsauctions - Ebay.com' 'TeeHee Socks' 'Stuff 4 Crafts'
 'drdealsazdrdealsaz - Ebay.com' 'Shade Tree Powersports - Walmart.com'
 'SupplyTiger - Walmart.com' 'Top Gear Outfitters'
 'allamerican2100 - Ebay.com' 'artofdeals'
 'Morrison Outdoor Power Equipment LLC.' 'Watch & Eyewear Center'
 'irena2015 - Ebay.com' 'GEM Products' 'Watching-Your-Time'
 'Stepani Style - Walmart.com' 'TazmasFashion' 'hc_huachen - Ebay.com'
 'Dancing Stone' 'Best Yet Collectables' 'Racketguys'
 'top-quality-fashion - Ebay.com' 'Art of Deals' 'wuyishang99 - Ebay.com'
 'simwood - Ebay.com' 'Focus Camera - Walmart.com'
 'set7apart2011 - Ebay.com' 'garson0808 - Ebay.com'
 'Peachy Gifts - Walmart.com' 'mbmerchandisembmerchandise - Ebay.com'
 'Full Throttle Outdoors LLC - Walmart.com' 'JewelryWe Jewelry'
 'Gen X Clothing - Walmart.com' 'Seven Times Six' 'WILLTOPINC'
 'lace_em_uplace_em_up - Ebay.com' 'augiejustinaugiejustin - Ebay.com'
 'Newegg' 'Outfitter Country - Walmart.com' 'Sell Lounge - Walmart.com'
 'Foot Paths Inc - Walmart.com' 'thezebrastand - Ebay.com'
 'Clothing Shop Online, LLC - Walmart.com'
 'SHIN BROTHERS JEWELERS - Walmart.com' 'CusCus'
 'ZappySales - Walmart.com' 'Fashion Group - Walmart.com'
 'Midwest Jewellery & Apparel' 'divas*delites - Ebay.com'

'pajama_pjc - Ebay.com' 'First Premier' 'TIJN'
 'Marcus Uniforms - Walmart.com' 'b2fitness_usa - Ebay.com' 'Vir Ventures'
 'Gem & Harmony'
 'Sacred Hoops and Hardwood Inc. dba Top 10 Sports - Walmart.com'
 'betterthebestbetterthebest - Ebay.com' "DIP 'N DIVE"
 'pescasseroli - Ebay.com' 'Lupo North America' 'trendsmax'
 'Avalanche Brands - Walmart.com' 'SHALL' 'The Displayer'
 'An Accessorized Woman' 'Summit Initiatives' "TK'sBargains"
 'Lionparts - Walmart.com' 'aSavings - Walmart.com' 'LBFEEL'
 'clickhere2shop' 'designersunglassesandwatches - Ebay.com' 'Almar'
 'AZ Sewing World' 'Luggage Pros - Walmart.com' 'Fuerza North America'
 'MaxiAids.com' 'taxesbybutlertaxesbybutler - Ebay.com' "Anna's Mall"
 'luxxe-shop - Ebay.com' 'crocophilestore - Ebay.com'
 'ecustomhitch - Walmart.com' 'Timepiece Trading LLC - Walmart.com'
 'quickstrikesportswearquickstrikesportswear - Ebay.com'
 'Sports By Sager Inc' 'Diabetes Health Supplies - Walmart.com'
 'Party & Costume SuperCentre' 'ShopeWatch Europe'
 'Greatlookz Inc - Walmart.com' 'Music123 - Walmart.com'
 'OJCommerce - Walmart.com' 'American Living Online - Walmart.com'
 'Forza Sports - Walmart.com' 'linyongshop - Ebay.com' 'Crazy4Tank'
 'eclectic-co.eclectic-co. - Ebay.com'
 'Superior Dollhouse Miniatures LLC - Walmart.com' 'loseee - Ebay.com'
 'lilystore2020 - Ebay.com' 'limitlessempire - Ebay.com'
 'shmad0215shmad0215 - Ebay.com' 'KONOV' 'eCentury LLC - Walmart.com'
 'Sports By Sager Inc. - Walmart.com' 'Rubicon Watch Company Limited'
 'GoldBar Watches' 'tourbon-safaristore - Ebay.com' 'Touch Enterprises'
 'fazamaOfazama0 - Ebay.com' 'scoresofmd - Ebay.com'
 'Sold by CellItAll and Fulfilled by Amazon.
 Gift-wrap available.'
 'CellItAll' 'The Jungle Store - Walmart.com' 'Tonzof - Walmart.com'
 'Fashion Traders' 'Discount Watch Store' 'Gaffos'
 'My Suit Habit - Walmart.com' 'Zillverzooom'
 'bauerjack1234bauerjack1234 - Ebay.com' 'topdawg_187 - Ebay.com'
 'Area Trend' 'Talin LLC - Walmart.com' 'West Coast Jewelry'
 'Fresh Threads Clothing' 'gzgogogocom - Ebay.com'
 'auntiej2007auntiej2007 - Ebay.com'
 'Little Family Members - Walmart.com'
 'teemzoneaowangteemzoneaowang - Ebay.com' 'teemzoneaowang - Ebay.com'
 'JustCalculators' 'Virtual Exchanges, Inc.'
 'frugalooiefrugalooie - Ebay.com' 'AvidMax - Walmart.com'
 'Wireless4Wireless' 'consignment305consignment305 - Ebay.com'
 'djhlywd2012djhlywd2012 - Ebay.com' 'Sunset Accessories' 'ST34' 'TXJ'
 'Purchase Pals LLC - Walmart.com' 'd-w-discountd-w-discount - Ebay.com'
 'Rembrandt Jewellers' 'US Trade Ent.' 'Back40Trading' 'Layger'
 'BIC Warehouse - (Includes all tax and Duties)' 'dealkings'
 'Batters Choice Sporting Goods - Walmart.com' 'buydisneystuff - Ebay.com'
 'Dave 'n' Mark's' 'daveo777 - Ebay.com' 'Killer Inc. - Walmart.com'
 'second_chance_shoes_zapatossecond_chance_shoes_zapatos - Ebay.com'


```
'TheAspecs' 'Hayneedle - Walmart.com' 'tomtop_w - Ebay.com'
'WatchesEmporium' 'Diamond Wish' 'Mad About Heels - Walmart.com'
'Marvellous Group' 'CostumeVille - Walmart.com'
'saisafashionsaisafashion - Ebay.com' 'EyeSave' 'ACI USA'
'authenticmenswearsauthenticmenswears - Ebay.com' 'A2ZID'
'ckbargainsretailckbargainsretail - Ebay.com' 'Mt.FUJI'
'wfd-sports - Ebay.com' 'ashford.com' 'shopeyeworld - Ebay.com'
'onemorebass - Ebay.com' 'glsne - Ebay.com' 'The Hat Pros'
'Cali Style Shop' 'Advance Buy' 'SUNTIMEZ' 'eliteonline2014 - Ebay.com'
'Converse' 'tackle-monster2015 - Ebay.com' 'VNOX Jewelry'
'fargoducksfargoducks - Ebay.com' 'XiaoXiao Ji'
'buzamare27buzamare27 - Ebay.com' 'SondoongMart.com'
'swissaltboxswissaltbox - Ebay.com' 'fashion-district - Ebay.com'
'dealsonlywebstore - Ebay.com' 'Deals Only - Walmart.com'
'weichangshun3000 - Ebay.com' 'Pens N More - Walmart.com'
'MyOfficeInnovations - Walmart.com' 'AMPM24CA'
'DEVINE MEDICAL - Walmart.com' 'keith264 - Ebay.com' 'Maciag Offroad'
'CosmeticMall - Walmart.com' 'Victory Tailgate' 'STAND UP Ranchers'
'Regal Brands LLC - Walmart.com' 'Gadsden and Culpeper'
'EyeSave - Walmart.com' 'Real Basics' 'My First Aid Company'
'sneakermanid - Ebay.com' 'Deals Outlet'
'Premium Overstock LLC - Walmart.com' 'Beston' 'Eyewearking'
'chadepetersonchadepeterson - Ebay.com' 'eBags - Walmart.com'
'Thekickshop' 'Legendary Whitetails - Walmart.com'
'The Deal Rack, LLC - Walmart.com' 'sportsdealer11 - Ebay.com'
'Deals Only Web Store' 'm3th1cal - Ebay.com'
'Dr. Sheepskin, Inc. - Walmart.com' 'SELECT ELITE' 'IM Survival'
'CPD Outlets' 'bbgolfstyle' 'Ithaca Sports - Walmart.com'
'bitbeatbitbeat - Ebay.com' 'Sundial Watches' 'All Authentic' 'Bikewagon'
'lanceman02lanceman02 - Ebay.com' 'Sports Watches Store'
'ace4bartace4bart - Ebay.com' 'Hurricane Golf - Walmart.com'
'BODYA Jewelry(Ships From China)' 'Select Footwear Online' 'eSource'
'Cuffcrazy LLC' 'Seek Optics' 'BikeSomeWhere' 'Jolia, LLC - Walmart.com']
```

Uniques: 748

Nans: 5516

5.6 Cleaning prices.condition column

```
[36]: get_unique_nans('prices.condition', df)
```

```
[nan 'new' 'New with tags' 'New with box' 'New without box' 'Pre-owned'
'New' 'New without tags' 'Brand New' 'New with defects' 'Used']
```

Uniques: 11

Nans: 6684

5.7 Cleaning colors column

This is fairly complex column, with a lot of messy colors. I will simply make it lower-case and remove dashes for the time being.

```
[37]: unique_colors, nans_color = get_unique_nans('colors', df, display=False)
unique_colors, nans_color

[nan
 'Red,Black,Sketchy Slant,Yellow,BlackBlue,BlackRed,Sharky Tribe,Pro
 Ocean,BlackBlueYellow,Gray,Helms,BlackGreenYellow,Blue,Green,Black-White,Black
 Multi,GrayBlack'
 'Brown' ... 'Mahogany Mahogany'
 'TangerineBlackGriffin,Griffin,AquaBlackGriffin,BlackBlackWhite,GriffinBlackWhi
 te'
 'Bay Brown']
Uniques: 2235
Nans: 8298

[37]: (array([nan,
              'Red,Black,Sketchy Slant,Yellow,BlackBlue,BlackRed,Sharky Tribe,Pro
 Ocean,BlackBlueYellow,Gray,Helms,BlackGreenYellow,Blue,Green,Black-White,Black
 Multi,GrayBlack',
              'Brown', ..., 'Mahogany Mahogany',
              'TangerineBlackGriffin,Griffin,AquaBlackGriffin,BlackBlackWhite,GriffinB
 lackWhite',
              'Bay Brown'], dtype=object), 8298)

[38]: df['colors'] = df['colors'].apply(lambda x: x.lower().replace(' ', '').
    ↳replace('-', '') if not pd.isnull(x) else x)
```

5.8 Cleaning manufacturer column

This one is ready for encoding as a categorical feature.

```
[39]: get_unique_nans('manufacturer', df)

[nan 'Maui And Sons' 'Twisted X' 'Puma' 'Filament' 'SOFIAMORE'
 'Robert Wayne' 'Roper' 'Puentes Denver' 'Elan Polo' 'Sebago'
 'SoftScience' 'Dr. Martens' 'Lanvin' 'Realtree'
 'Quicksilver Wholesale Inc.' 'Kenneth Cole' 'Nike' 'Stacy Adams' 'Oakley'
 'Genuine Grip' 'El Naturalista' 'New Balance' 'STC' 'RBX'
 'New Balance Numeric' 'Norcross Safety Prod' 'PRIMESOURCE, INC.'
 'Converse' 'Kenneth Cole Reaction' 'DVS' 'Qianyuyu' 'Keuka SureGrip'
 'Old Friend' 'Skechers' 'Isotoner' 'Tretorn' 'Metro Jewelry' 'TOMS'
 'Onguard' 'Tommy Hilfiger' 'Toms' 'Hush Puppies' 'Kenneth Cole NY'
 'GameDay Boots' 'Georgia Boot' 'Rider Mens Footwear' 'Polo Ralph Lauren'
 'Massimo Matteo' 'Hunter' 'Bates' 'East Majik' 'Project Emotion' 'Cashel'
 'Smart Step' 'VANS' 'Propet' 'Mizuno' 'WOLVERINE WORLDWIDE']
```

'Kylin Express' 'Gordon Rush' 'Gameday Boots' 'Work America'
 'Forever Collectibles' 'Reebok' 'Nocona' 'POV ACTION VIDEO CAMERAS'
 'Casio' 'Timberland' 'Ralph Lauren' 'Florsheim' 'Bedroom Athletics'
 'Perry Ellis' 'Bar III' 'Muk Luks' 'VALUE BRAND'
 'Onitsuka Tiger by Asics' 'Boss Hugo Boss' 'Baffin Inc' 'Deckers'
 'Vision Street Wear' 'Calcutta' 'Eastland' 'Daxx' 'Dickies'
 'Forum Novelties' 'Quiksilver' 'PUMA SAFETY SHOES' 'TONY LAMA BOOT CO.'
 'Reef' 'City Hunter' 'Original SWAT' 'Coronado' 'ROCKY' 'Merrell' 'Sanuk'
 '3N2, LLC' 'Crocs' 'saucony' 'Chippewa' 'Bearpaw' 'Rubies' 'Wishpets'
 'SERVUS BY HONEYWELL' 'The North Face' 'ELLIE SHOES' 'SHOES FOR CREWS'
 'Wembley' 'La Sportiva' 'Cushe' 'Zoot' 'Newton Running' 'Baffin'
 'Mercanti Fiorentini' 'Generic Surplus' 'Under Armour' 'Ridge Footwear'
 'DIAMONDBACK' 'TAG Heuer' 'Co-operative Feed Dealers Inc.' 'Clarks'
 'LaCrosse Footwear' 'Sperry Top Sider' 'Minnetonka' 'Scarpa' 'Gola'
 'Timex' '5150' 'Rocky' 'Jordan' '29 Porter Rd' 'Bata Shoe'
 'Principle Plastics Inc' 'Rctech' 'Hi-Tec' 'John Varvatos' 'Seavees'
 'PleaserUSA' 'Nobles Health Care Products' 'H.S. Trask'
 'Diesel Black Gold' 'Empire' 'Justin' 'L.B. Evans' 'Deluxe Comfort'
 'Bogs' 'Comfy Feet LLC' 'My Team' 'Keen' 'RVCA' 'Exustar'
 'Pacific Shoe Corp.' 'DAN POST' 'Birkenstock' 'Chevron Pattern Watch'
 'Forever' 'Red Wing Brands' 'Tingley' 'Air Jordan' 'Columbia' 'Rothco'
 'SCOTT' 'Grendene' 'Dockers' 'TINGLEY RUBBER' 'WORK MASTER'
 'Weatherproof' 'Montrail' 'AIR JORDAN' 'Ariat' 'Wolverine World Wide'
 'Frye' 'Levi's' 'Lacoste' 'Northside' 'Emerica' 'Fallen' 'Chocolat Blu'
 'Mark Nason' 'NORCROSS SAFETY' 'Hey Dude' 'Mil-Spec Plus' 'Helly Hansen'
 'Team Beans' 'Tru-Spec' 'Charles By Charles David'
 'THE ORIGINAL MUCK BOOT CO.' '2(x)ist' 'J'S AWAKE' 'Dan Post'
 'HXW040004110AB35012' 'Forever Collectible' 'JUSTIN ORIGINAL WORKBOOTS'
 'Vans Inc' 'Adidas' 'ERGOMATES' 'REEF' 'Costa Del Mar' 'DDI' 'Bokos'
 'Fila' 'Yaleet Inc.' 'Lamo' 'Vans Footwear' 'BC Footwear' 'J75 by Jump'
 'Saucony' 'Wells Lamont' 'Diamondback' 'ROYER' 'Warson Brands'
 'Dearfoams' 'Danner' 'Marc Ecko Cut & Sew' 'Neff' 'Morris Costumes'
 'Alfani' 'Trevco' 'Rockport' 'Ferrini' 'Rasta Imposta' 'The Generic Man'
 'OOFOS' 'Durango' 'Deer Stags' 'Unlisted Kenneth Cole' 'Hartz' 'Laredo'
 'SCRUB ZONE' 'Itasca' 'Avia' 'Daniel Wellington' 'Paul Green'
 'Giorgio Brutini' 'Lacrosse Footwear Inc.' 'Viking' 'K-Swiss' 'XTRATUF'
 'Telic' 'Sloggers' '3N2' 'asics' 'Buckle Rage' 'Creative Recreation'
 'Carrera' 'Boreal' 'DuPont' 'Dolomite' 'Crevo' 'Cole Haan'
 'Dr. Martens Air Wair' 'Caterpillar' 'Gola Classics' 'Old West'
 'London Fog' 'O'Neill' 'CALLAWAY' 'BOSTONIAN' 'Bruno Magli'
 'Johnston & Murphy' 'O'Neal' 'Romeo and Juliette Inc.'
 'Fruit of the Loom' 'Asics' 'Stetson' 'Slippers International'
 'RANGER BY HONEYWELL' 'Dvs' 'Totes' 'Kamik' 'Pajar'
 'Checkered Pattern Watch' 'unknown' 'teva' 'Icebug' 'Swear London'
 'Reliable of Milwauakee' 'Tin Haul' 'SilverHooks' 'Tesh' 'Robert Graham'
 'MUCK BOOTS' 'Dan Post Boots' 'Portland Product Werks' 'Alexanders'
 'Aldo' 'Timberland Pro' 'Palladium' 'Swear' 'Comfy Feet' 'Haggar'
 'Dikamar' 'Stafford' 'Chinook' 'Justin Boots' 'Bed Stu' 'MuckBoot'

'CIRCA' 'Sorel' 'Wolverine' 'VIONIC BY ORTHAHEEL' 'Belvedere Inc'
 'Emporio Armani' 'ASICS' 'Pleasers USA, Inc.' 'Tod's'
 'Rubies Costume Co. Inc' 'Calvin Klein' 'Bheema' 'HEAD' 'Etnies'
 'Leg Avenue' 'Andrew Marc' 'Custom Leathercraft' 'Red Wing' 'Blanco'
 'Reef Footwear' 'Natural World' 'RENAISSANC' 'Rugged Shark' 'Safilo'
 'Grenade' 'Reliable of Milwaukee' 'Mintcraft' 'Stabilicers' 'Mephisto'
 'Hawaiian Classics' 'Silver Lilly' 'Wood N Stream' 'Alpinestars' 'Born'
 'Lugz' 'MukLuks' 'Dije California' 'Acacia' 'Dunlop' 'Alessandro Monelli'
 'Dr. Scholl's' 'Columbia Sportswear Company' 'HideAways by L.B. Evans'
 'VIKING' 'Rainbow' 'Servus' 'Elope' 'Vans Skate' 'Online Stores' 'Tecs'
 'Harbor Footwear' 'Minnetonka Moccasin Company, Inc.'
 'Kenneth Cole New York Footwear Mens' 'Ray-Ban' 'Reebok Footwear'
 'Noble Outfitters' 'adidas' 'Luccchese' 'Bushnell' 'Gucci' 'Icon'
 'Green Mountain Imports' 'BOC MENS' 'Onitsuka Tiger' 'Weaver Leather'
 'Cat Footwear' 'Testosterone Shoes' 'Genfoot America' 'Ecco' 'Boulet'
 'Acorn' 'Hoka One One' 'Rocky Brands' 'Ugg Australia'
 'Maison Martin Margiela' 'Steve Madden' '1883 by Wolverine' 'CHIPPEWA'
 'Aston Grey Collection' 'ONGUARD' 'MEZLAN' 'BOSS MFG COMPANY'
 'Spring Footwear' 'Tommy Bahama Group Inc.' 'Ellie Shoes Inc'
 'Klone Lab LLC' 'Comic Images' 'Cycleur de Luxe' 'Rod Patrick' 'Firefly'
 'Tony Lama' 'IRON AGE' 'RJ's Fuzzies' 'HUF Men's Footwear' 'Flojos'
 'SOLE Footwear' 'Sperry' 'Nautica' 'Studio Ink' 'Sendra Boots'
 'Honeywell' 'SCARPA' 'AGS Footwear' 'Major Surplus' 'ROCKPORT' 'Olukai'
 'Gravity Defyer' 'BioWorld' 'DiamondBack' 'Vionic' 'Cudas Footwear' 'GBX'
 'Circa' 'Stuhrling Original' 'Cat' 'Unassigned' 'Modern Vintage'
 'GE Animation' 'RG Costumes' 'USA Dawgs, Inc.' 'And1' 'Lakai'
 'Puma Safety Shoes' 'Izod' 'Winthrop' 'Rush by Gordon Rush'
 'US Polo Assn' 'Kenetrek' 'Hallmark' 'Chrome' 'Oofos' 'Huf' 'Old Gringo'
 'Original Penguin' 'Broken T' 'Voodoo Tactical' 'Vito Rossi'
 'Original S.W.A.T' 'OluKai' 'Gorila' 'Blackhawk' 'Chums' 'USA Dawgs'
 'Tom Ford' '5.11 Tactical' 'Ferrari' 'Carhartt' 'Gardena' 'Zanzara'
 'Complete Medical Supplies' 'Lowa' 'Norcross Safety' 'John Deere'
 'P.W. Minor' 'Invicta' 'PF Flyers' 'Brunello Cucinelli' 'PF-Flyers'
 'Liberty Glove & Safety' 'Superior Glove Works' 'Alwin'
 'Kaz Home Environment' 'FitFlop' 'Supra' 'New Balance Athletic Shoe Inc.'
 'CTM' 'Hugo Boss' 'Gianvito Rossi' 'Smiffy's' 'Croakies'
 'South China Imports' 'Spira' 'Nike SB' 'Toms Footwear' 'Saint Laurent'
 'Principle Plastics' 'Bacco Bucci' 'Marvel' 'Masimo New York'
 'Alexander McQueen By Puma' 'Ben Sherman' 'Command'
 'RG Barry Corporation' 'MADE Cam Newton' 'HMK' 'Joe's' 'Rycore' 'Drew'
 'FunWorld' 'DrMartens' 'Moods of Norway' 'Marcolin' 'Dsquared2' 'Diesel'
 'GOLDEN RETRIEVER OUTDOOR FOOTWEAR' 'Redback Boots' 'Alexander'
 'Timberland Earthkeepers' 'CKX' 'Marc Joseph' 'PUMA' 'Speedo' 'Franco'
 'SPERRY' 'JewelryWe' 'Fairly Odd Novelties' 'Mik' 'Bugatchi' 'Sunville'
 'Ranger' 'Oliberte' 'Skechers Work' 'Rugged Country'
 'Red Label By Giovanni' 'Rolson' 'Emu' 'Hadley-Roma' 'Seiko'
 'McRae Industrial' 'Magliner' 'Dragon Watch' 'CLARKS' 'Jed North'
 'Woolrich' 'Tommy Bahama' 'Professional's Choice' 'Osiris Shoes' 'Buks'

'Headsweats' 'Birch Tree Watch' 'Proline' 'Mizuno Running USA'
 'Performance Tool' 'Sea Turtle Watch' 'Inov-8'
 'Qingdao Gloria International Co Ltd' 'DS Work' 'WINTER WALKING'
 'Independent Boot Company' 'BEKINA' 'Mancini' 'Slipperooz by Deer Stags'
 'RW by Robert Wayne' 'Y-3' 'Treds' 'J Shoes' 'Trask' 'Supplier Generic'
 'Outdoor Recreation Group' 'Cocoons' 'Ironlace' 'Jobst' 'Lucien Piccard'
 'Forum Novelties Inc.' 'Zoot Sports' 'Greatlookz Fashion' "Levy's"
 'Dipsters' 'Venum' 'Marc Fisher' 'Lacrosse' 'Adan' 'Khombu'
 'Robinson Outdoor Products' 'West Coast Jewelry'
 'Marc New York By Andrew Marc' 'Grendene USA Inc.' 'Eastland Shoe'
 'LACROSSE' 'Made Cam Newton' 'SANUK' 'KC Caps' 'P&W New York'
 'Blundstone' 'Serengeti Eyewear' 'Lorus' 'Armitron'
 'Franco American Novelties' 'Cufflinks Inc' 'SO & CO New York'
 'British Knights' 'Berne' 'A|X Armani Exchange' 'Fred Perry' 'A. Testoni'
 'SoftWalk' 'G.H. Bass & Co.' 'Rockport XCS' 'Gold Toe' 'Muck'
 'Vasque Footwear' 'SunCloud' 'Mechanix Wear' 'Guess' 'Amer Sports'
 'U.S Polo Assn.' 'Powerbank2013' 'AmeriLeather' 'Okabashi' 'Corral'
 'Tsubo' 'Sorel Footwear' 'HyTest by Wolverine' 'NEOS OVERSHOE' 'Smoky'
 'Timeless Toys Inc' 'Armani Jeans' 'MAH' 'Cherry Blossom'
 'Allrounder By Mephisto' 'NATIONAL SAFETY APPAREL' 'MUCK BOOT'
 'OTZ Shoes Inc' 'totes Isotoner' 'Pearl Izumi' 'Faconnable' 'Darn Tough'
 'Tissot' 'Jimmy Choo' 'Double H' 'buytra' 'MCM' 'Ahnu'
 'Resolve By RW Footwear' 'Faded Glory' 'Do.BOMRVII' 'Vince Camuto'
 'MICROFLEX' 'True Religion' 'Moneysworth & Best' 'Array'
 'Sperry Top-Sider' 'Elten' 'DeRoyal' 'Five Ten'
 'Norcross Safety Products' 'Mason Shoe' 'Aircraft Watch' 'Wilson'
 'Guide Gear' 'Gadsden and Culpeper' 'COSMAS JAVA' 'Rogue' 'Ugg'
 'Latitudes Inc.' 'Pleaser' 'LeDonne Leather' 'UGG MENS' 'Donald J Pliner'
 'Suncloud' 'Legendary Whitetails' 'Outdoor Cap' 'Marc New York' 'Biion'
 'Swiss Legend' 'Polo' 'Kickers' 'Vasyli' 'Urban Pipeline' 'Easy USA'
 'Salomon' 'BODYA' 'MF Western' 'Luxottica' 'Minav Corp.'
 'Rainbow Sandals' 'Muck Boots' 'AH by Android Homme' 'Rubies Costume Co'
 'Vasque' 'POLO Ralph Lauren']

Uniques: 668

Nans: 12646

5.9 Cleaning prices.offer column

I will leave it as is.

```
[40]: get_unique_nans('prices.offer', df)
```

```
['REDUCED USD 12.10' nan 'SAVINGS_AMT USD 9.80' ... '5% off' '4% off'
 'CLEARANCE USD 70.02']
```

Uniques: 1237

Nans: 13273

5.10 Cleaning prices.shipping column

I will broadly categorize this variable into “Free” vs “Not Free”

```
[41]: get_unique_nans('prices.shipping', df)
```

```
[nan 'FREE' 'AUD 25.00' 'free' 'USD 10.99' 'AUD 14.99' 'USD 8.99'
'USD 5.09 shipping' 'FREE Returns on some sizes and colors.' 'USD 7.58'
'AUD 35.00' 'USD 10.00' 'FREE shipping on orders 50 +'
'FREE Shipping. FREE Returns.' 'USD 8.99 shipping' 'USD 7.99' 'GBP 12.95'
'USD 9.99' 'USD 5.99' 'AUD 12.42' 'USD 3.34' 'USD 2.74' 'USD 4.99'
'USD 3.02' 'USD 6.95' 'FREE Shipping' 'USD 0.01' 'USD 7.95'
'FREE Returns.' 'USD 3.99' 'AUD 19.95' 'AUD 10.00' 'AUD 24.95'
'USD 17.00' 'USD 9.95' 'USD 4.35' 'AUD 24.00'
'FREE shipping available on 50 + orders'
'FREE SHIPPING ON ORDERS OVER $125. $5 FLAT-RATE ON OTHER ORDERS $125 AND LESS'
'USD 6.99' 'USD 18.00' 'GBP 14.99(approx. US $19.58)US $19.58'
'GBP 14.99' 'USD 25.00 shipping' 'USD 6.80' 'USD 13.99' 'USD 2.99'
'USD 3.95' 'USD 6.45' 'USD 5.00 shipping' 'USD 9.00' 'USD 24.95'
'USD 13.90' 'USD 6.50' 'USD 12.99' 'USD 8.65' 'AUD 8.99' 'USD 20.00'
'USD 4.49 shipping' 'USD 7.00' 'USD 6.00' 'USD 35.00' 'USD 7.92'
'USD 6.92' 'USD 18.90' 'FREE shipping on orders 35 +' 'AUD 47.50'
'USD 6.90 shipping' 'USD 2.95' 'USD 12.65' 'USD 13.45' 'GBP 20.00'
'AUD 46.17' 'USD 9.17' 'USD 4.95' 'USD 8.95'
'Always free shipping, fittings & returns' 'CAD 8.99' 'USD 5.90'
'USD 30.00' 'USD 12.00' 'AUD 22.00' 'GBP 8.00' 'USD 15.00' 'USD 14.99'
'USD 17.54' 'USD 19.23' 'USD 10.53' 'FREE One-Day Shipping Free Returns'
'USD 6.99 shipping' 'USD 7.50' 'USD 2.50 shipping' 'USD 3.78' 'GBP 3.00'
'FREE Shipping on orders over USD 35.00 FREE Returns.' 'USD 3.00'
'USD 8.90' 'USD 25.00' 'USD 5.00' 'USD 17.99' 'USD 2.45 shipping'
'USD 8.85' 'USD 5.74' 'USD 3.99 shipping' 'AUD 19.99' 'AUD 39.70'
'AUD 49.95' 'USD 32.00' 'AUD 20.00'
'FREE Shipping on orders over USD 35, FREE Returns.' 'USD 8.50'
'USD 11.75' 'USD 4.00' 'USD 5.95' 'USD 24.50' 'USD 6.35' 'USD 8.58'
'USD 7.85' 'AUD 21.10' 'USD 14.95'
'You want your order delivered quickly and safely. All our orders are shipped
with USPS and delivery is free.'
'USD 8.00' 'USD 15.25' 'USD 10.65' 'USD 15.90'
'FREE Shipping on orders over USD 35.00' 'USD 6.10' 'USD 4.63' 'USD 9.60'
'USD 11.66' 'USD 2.45' 'USD 5.95 shipping' 'USD 7.59' 'FREE Shipping.'
'USD 1.69' 'USD 2.00 shipping' 'USD 6.11'
'FREE Returns on some sizes and colors. Details'
'FREE Shipping on orders over USD 49.00 FREE Returns.'
'FREE Shipping on eligible orders. FREE Returns.' 'USD 45.00' 'USD 7.65'
'AUD 9.99' 'AUD 29.00' 'USD 11.42 shipping' 'USD 13.00' 'USD 11.00'
'USD 5.79 shipping' 'USD 2.75' 'USD 40.00' 'FREE shipping on orders 50+'
'USD 3.49' 'USD 5.60' 'USD 9.58 shipping' 'USD 4.50 shipping' 'CAD 10.23'
'USD 9.53' 'EUR 2.59' 'AUD 37.00' 'USD 11.60' 'USD 21.40' 'USD 4.14'
'USD 24.00' 'USD 11.95 shipping' 'GBP 12.00'
```

```
'FREE Shipping on orders over USD 49.00' 'AUD 40.00' 'USD 3.80 shipping'
'USD 19.99' 'USD 13.60' 'USD 12.90' 'GBP 18.00' 'USD 1.99'
'USD 4.99 shipping' 'USD 3.94' 'AUD 49.95(approx. US $38.05)US $38.05'
'CAD 4.68' 'CAD 5.04' 'USD 4.45' 'USD 5.50' 'USD 3.50' 'USD 13.95'
'Free Shipping for Prime Members Fast, FREE Shipping with Amazon Prime'
'AUD 39.95' 'USD 3.75' 'USD 12.50' 'USD 10.49' 'USD 11.99'
'FREE Shipping on eligible orders.' 'GBP 10.99' 'USD 31.00' 'USD 21.04'
'CAD 19.95' 'AUD 7.99' 'USD 3.40' 'USD 7.89' 'USD 3.98' 'USD 3.52'
'EUR 0.90' 'USD 9.50' 'USD 4.50' 'Free Shipping' 'USD 13.40 shipping'
'USD 29.00' 'USD 11.20' 'USD 11.88' 'CAD 5.00' 'USD 15.00 shipping'
'CAD 5.54' 'USD 7.33 shipping' 'USD 12.95' 'USD 3.25' 'USD 7.76'
'USD 6.48' 'AUD 29.99' 'CAD 3.99' 'CAD 6.90' 'USD 15.95' 'USD 4.84'
'USD 4.89 shipping' 'AUD 29.95' 'USD 9.99 shipping'
'Free Shipping for Prime Members Fast, FREE Shipping with Amazon Prime. FREE
Returns.'
```

```
'USD 16.00' 'CAD 8.75' 'USD 21.00' 'USD 1.29' 'USD 3.86' 'USD 5.25'
'CAD 5.99' 'USD 28.00' 'USD 5.75' 'USD 0.99' 'USD 9.90' 'USD 19.00'
'CAD 8.00' 'CAD 7.00' 'USD 6.15 shipping' 'CAD 11.38' 'USD 10.60'
'USD 2.97 shipping' 'USD 14.17' 'USD 7.90' 'USD 7.95 shipping'
'FREE pickup' 'CAD 30.78' 'CAD 11.99' 'USD 42.00' 'USD 50.00' 'USD 11.50'
'USD 10.71' 'USD 14.50' 'AUD 36.70' 'AUD 28.71' 'USD 12.20' 'USD 13.65'
'USD 18.99' 'USD 19.50' 'USD 4.16' 'USD 4.90 shipping' 'USD 9.55'
'GBP 10.00' 'USD 23.00 shipping' 'USD 10.37' 'USD 5.49 shipping'
'USD 3.59' 'EUR 0.99' 'EUR 2.99' 'USD 6.49' 'USD 5.10' 'CAD 8.26'
'CAD 12.19' 'CAD 6.00' 'CAD 4.14' 'CAD 3.15' 'USD 12.75'
'USD 5.99 shipping' 'CAD 10.99' 'AUD 45.00' 'AUD 15.00'
'USD 19.35 shipping' 'USD 19.75 shipping' 'USD 26.35 shipping'
'GBP 12.60' 'GBP 2.95' 'CAD 6.95' 'USD 6.70' 'USD 36.99' 'USD 27.50'
'CAD 7.77' 'AUD 29.60' 'USD 6.75' 'USD 2.29' 'USD 3.54' 'USD 16.95'
'USD 14.95 shipping' 'USD 14.55 shipping' 'USD 5.45' 'USD 4.34'
'USD 7.15' 'USD 22.00' 'AUD 1.15' 'CAD 20.24' 'USD 12.45'
'USD 10.18 shipping' 'USD 10.98' 'USD 13.98 shipping' 'USD 7.50 shipping'
'USD 2.50' 'USD 9.86' 'EUR 15.00' 'USD 4.49' 'GBP 15.00' 'CAD 28.99'
'CAD 9.05' 'USD 15.17' 'EUR 0.69']
```

Uniques: 317

Nans: 13568

```
[42]: def process_shipping(x):
        if pd.isnull(x):
            return x

        x = x.lower()
        if 'free' in x:
            return 'free'
        else:
            return 'not free'
        return x
```

```
df['prices.shipping'] = df['prices.shipping'].apply(process_shipping)
df['prices.shipping'].value_counts()
```

```
[42]: free          4049
      not free     1677
      Name: prices.shipping, dtype: int64
```

5.11 Cleaning sizes column

To simplify, I will calculate how many sizes are available for a particulate shoe.

```
[43]: get_unique_nans('sizes', df)
```

```
[nan 'XL' '10,10.5,11,11.5,12,13,8,8.5,9,9.5' ... '5.5,7,6,6.5,4.5'
 '10.5,10,11.5,14' '8.5 W,7 W']
Uniques: 1088
Nans: 13326
```

```
[44]: def process_sizes(x):
      if pd.isnull(x):
          return x
      return len(x.split(','))

df['sizes'] = df['sizes'].apply(process_sizes)
df['sizes'].value_counts()
```

```
[44]: 1.0      2703
      2.0      626
      3.0      541
      4.0      325
      5.0      286
      6.0      247
      7.0      195
      10.0     179
      8.0      152
      9.0      144
      12.0     121
      11.0      98
      13.0      91
      15.0      83
      14.0      18
      26.0      16
      29.0      15
      19.0      14
      18.0      12
      17.0       9
      34.0       9
```



```

49.0      7
20.0      7
16.0      6
46.0      6
30.0      6
28.0      6
60.0      6
56.0      5
27.0      5
43.0      4
22.0      4
21.0      4
23.0      3
32.0      3
44.0      3
53.0      3
42.0      2
25.0      2
24.0      1
31.0      1
Name: sizes, dtype: int64

```

5.12 Cleaning dimension column

Let's see if dimension can be easily processed.

To simplify, I will create a dimension number as the sum of all 3 dimensions. This will help to deal with inconsistencies of the order of the three dimensions (e.g. is it height-width-depth or depth-height-width or some other combination).

```
[45]: unique_dims, _ = get_unique_nans('dimension', df, display=False)
```

```

[nan '10.5 in x 11.25 in x 1.2 in' '15 in x 17 in x 5 in'
 '1 in x 1 in x 1 in' '6 in x 3 in x 3 in' '8 in x 12 in x 3 in'
 '1 in x 1 in x 10 in' '1 in x 3 in x 10 in' '1 in x 1 in x 3 in'
 '10 in x 10 in x 10 in' '1 in x 5 in x 10 in' '12 in x 12 in x 4 in'
 '14.0 in x 11.5 in x 2.0 in' '2 in x 3 in x 1 in'
 '7.51 in x 10.23 in x 1.32 in' '12 in x 13 in x 4 in'
 '5 in x 10 in x 3 in' '3.94 in x 5.91 in x 1.97 in' '1 in x 2 in x 10 in'
 '6 in x 6 in x 6 in' '5 in x 6 in x 4 in' '0.01 in x 0.01 in x 0.01 in'
 '2.75 in x 4.0 in x 10.625 in' '1 in x 6 in x 6 in'
 '7.6 in x 9.5 in x 1.4 in' '13.0 in x 7.0 in x 5.0 in'
 '5 in x 2 in x 5 in' '10 in x 15 in x 4 in' '3 in x 3 in x 2 in'
 '5 in x 7.4 in x 1.5 in' '17 in x 21 in x 5 in'
 '6.0 in x 6.0 in x 1.0 in' '1 in x 5 in x 13 in' '4 in x 8 in x 8 in'
 '12.4 in x 16.1 in x 8.4 in' '1 in x 5 in x 11 in'
 '7.5 in x 8.6 in x 3.5 in' '4 in x 13 in x 1 in' '7 in x 10 in x 1 in'
 '17 in x 23.63 in x 7.09 in' '1 in x 5 in x 12 in'
 '18.5 in x 11.81 in x 2.76 in' '13 in x 26 in x 2 in'

```

'17.75 in x 27 in x 1 in' '2 in x 6 in x 10 in'
 '17.5 in x 27 in x 1.06 in' '7.9 in x 11.7 in x 4.4 in'
 '10 in x 5 in x 15 in' '6.5 in x 5.5 in x 4.5 in' '6 in x 2 in x 3 in'
 '1 in x 5 in x 14 in' '8 in x 12 in x 8 in' '14 in x 24 in x 4 in'
 '5.0 in x 13.0 in x 21.0 in' '10.5 in x 12.5 in x 15.5 in'
 '7 in x 2 in x 3 in' '4 in x 4 in x 5 in' '5 in x 7 in x 12 in'
 '4 in x 8 in x 2 in' '14 in x 13 in x 15 in' '7.5 in x 6.5 in x 8.5 in'
 '1 in x 11 in x 0.25 in' '30 in x 30 in x 30 in' '1 in x 7 in x 5 in'
 '19 in x 14 in x 0.01 in' '18.9 in x 14.96 in x 12.6 in'
 '4 in x 6 in x 6 in' '5 in x 10 in x 10 in' '16.0 in x 15.0 in x 20.0 in'
 '10.3 in x 12.1 in x 6 in' '5 in x 4 in x 6 in' '1 in x 1 in x 9 in'
 '1 in x 4 in x 4 in' '1 in x 4 in x 10 in' '18 in x 24 in x 4 in'
 '17 in x 17.5 in x 1.13 in' '11 in x 16 in x 10 in' '8 in x 2 in x 9 in'
 '13 in x 29 in x 5 in' '0.1 in x 0.1 in x 0.1 in' '4 in x 6 in x 4 in'
 '15 in x 14 in x 16 in' '9 in x 2 in x 7 in' '6 in x 10 in x 13 in'
 '5.9 in x 10.4 in x 0.7 in' '6 in x 6 in x 1 in'
 '13.0 in x 7.5 in x 4.5 in' '11.02 in x 1.89 in x 4.41 in'
 '11.5 in x 11.5 in x 11.5 in' '8 in x 10 in x 6 in'
 '3 in x 6 in x 0.25 in' '17.32 in x 6.69 in x 1 in'
 '16.0 in x 13.0 in x 3.0 in' '9.3 in x 11 in x 5 in'
 '6.1 in x 10.4 in x 5.4 in' '2.6 in x 4.6 in x 14.95 in'
 '23 in x 22 in x 24 in' '3 in x 3 in x 1 in' '4.5 in x 10.8 in x 1.6 in'
 '2.64 in x 2.5 in x 2.4 in' '7.2 in x 9.6 in x 1.5 in'
 '5 in x 12 in x 6 in' '17.0 in x 7.0 in x 48.0 in' '4 in x 9 in x 1 in'
 '11.0 in x 4.125 in x 3.75 in' '3.9 in x 5.46 in x 0.98 in'
 '1 in x 1 in x 6 in' '7.48 in x 10.08 in x 0.79 in'
 '12 in x 11 in x 13 in' '8.7 in x 10.5 in x 2.6 in' '5 in x 12 in x 2 in'
 '16.0 in x 11.0 in x 3.0 in' '5.5 in x 6.5 in x 7.5 in'
 '9.84 in x 12.2 in x 0.39 in' '6.3 in x 14 in x 1.5 in'
 '10.5 in x 16 in x 2 in' '14.4 in x 12.75 in x 9.0 in'
 '3.9 in x 5.46 in x 1.56 in' '2 in x 2 in x 1 in' '7 in x 9 in x 1 in'
 '14 in x 18 in x 4 in' '7.2 in x 16 in x 1.1 in'
 '17.5 in x 27 in x 1.43 in' '2 in x 10 in x 2 in'
 '10.5 in x 15.8 in x 0.9 in' '6.22 in x 12.44 in x 1.34 in'
 '13.0 in x 8.0 in x 5.0 in' '14 in x 17 in x 3 in'
 '11.7 in x 14.04 in x 0.2 in' '13 in x 18 in x 2 in'
 '14.01 in x 12.51 in x 9.51 in' '14 in x 12.5 in x 12.5 in'
 '9 in x 1 in x 4 in' '13 in x 16 in x 5 in'
 '7.64 in x 11.02 in x 0.79 in' '18 in x 9 in x 6 in'
 '9 in x 14 in x 5 in' '4 in x 5 in x 10 in' '7.4 in x 15.8 in x 1.3 in'
 '13.4 in x 15.7 in x 4.7 in' '11.0 in x 4.5 in x 3.5 in'
 '1.5 in x 12 in x 4.5 in' '9.5 in x 9.5 in x 9.5 in' '4 in x 3 in x 5 in'
 '6.2 in x 7.2 in x 1.5 in' '11 in x 5 in x 13 in'
 '9.6 in x 9.7 in x 9.2 in' '12 in x 14 in x 6 in' '12 in x 13 in x 22 in'
 '8.6 in x 10.6 in x 2.1 in' '6.9 in x 7.1 in x 1.1 in'
 '6.5 in x 5.5 in x 7.5 in' '5.1 in x 7.8 in x 1.8 in'
 '3 in x 3 in x 3 in' '13.25 in x 17 in x 6 in' '5.5 in x 7.5 in x 8.5 in'
 '14 in x 18 in x 6 in' '13.0 in x 8.0 in x 4.0 in' '1 in x 7 in x 9 in'

'4.0 in x 8.0 in x 10.75 in' '8.8 in x 11.8 in x 2.3 in'
 '1 in x 4 in x 7 in' '8 in x 14 in x 14 in' '1.97 in x 2.76 in x 0.39 in'
 '2 in x 2 in x 2 in' '0.5 in x 0.6 in x 6 in' '3.25 in x 3.5 in x 1 in'
 '6 in x 8.4 in x 1.4 in' '3.5 in x 7.75 in x 9 in'
 '22.0 in x 15.0 in x 15.0 in' '9 in x 5 in x 13 in'
 '5.0 in x 13.0 in x 7.0 in' '6.3 in x 10.4 in x 2 in'
 '9 in x 2 in x 6 in' '5 in x 9 in x 1 in' '1.97 in x 2.36 in x 1.97 in'
 '3 in x 7.6 in x 1.3 in' '8.9 in x 13.8 in x 5.2 in'
 '6.2 in x 7.6 in x 1.8 in' '12 in x 9 in x 0.75 in'
 '7.2 in x 11.3 in x 5.6 in' '4 in x 4 in x 7 in'
 '7.6 in x 11 in x 3.5 in' '6.4 in x 11.9 in x 4.9 in'
 '11.5 in x 14.3 in x 6.2 in' '5 in x 4 in x 9 in'
 '12.5 in x 13.5 in x 2.5 in' '14.0 in x 7.0 in x 5.0 in'
 '10 in x 10 in x 15 in' '6.69 in x 10.39 in x 2.28 in'
 '12 in x 5.5 in x 13.5 in' '4 in x 1 in x 9 in'
 '4.5 in x 7.25 in x 10.5 in' '7.56 in x 12.25 in x 7.56 in'
 '2.2 in x 23.2 in x 2 in' '6 in x 5 in x 10 in'
 '4.4 in x 8.7 in x 1.7 in' '3 in x 2 in x 5 in'
 '10.53 in x 12.09 in x 0.78 in' '10.8 in x 13.1 in x 8.5 in'
 '10 in x 12 in x 6 in' '12 in x 12 in x 12 in'
 '11.2 in x 15.2 in x 0.4 in' '5 in x 1 in x 9 in'
 '19.5 in x 14.0 in x 9.75 in' '4.1 in x 10.5 in x 1 in'
 '7.5 in x 4 in x 8.5 in' '5.2 in x 11 in x 3.3 in'
 '0.16 in x 1.1 in x 1.7 in' '18.5 in x 16.0 in x 18.5 in'
 '16 in x 15 in x 17 in' '4.5 in x 14 in x 3 in' '10 in x 14 in x 4 in'
 '1 in x 14 in x 1 in' '6 in x 13 in x 5.8 in' '4 in x 11 in x 5 in'
 '6.7 in x 12.3 in x 4.9 in' '7 in x 2 in x 9 in'
 '11 in x 11.9 in x 1.1 in' '6.5 in x 10.2 in x 3.3 in'
 '10.6 in x 14.4 in x 4.7 in' '8.0 in x 6.0 in x 0.5 in'
 '1 in x 6 in x 3 in' '14.5 in x 16.3 in x 2.2 in'
 '4.3 in x 8.8 in x 4.2 in' '7.4 in x 11.5 in x 4.4 in'
 '5.8 in x 11.4 in x 2.5 in' '6.22 in x 12.44 in x 1.1 in'
 '2 in x 3 in x 4 in' '11 in x 17 in x 10 in' '1.6 in x 1.5 in x 1.7 in'
 '8.2 in x 9.9 in x 1.7 in' '1 in x 7 in x 1 in' '1 in x 2 in x 1 in'
 '15 in x 15 in x 15 in' '9.76 in x 11.34 in x 1.02 in'
 '4.25 in x 13.0 in x 15.0 in' '5 in x 10 in x 11 in'
 '13 in x 16.5 in x 11.7 in' '7.7 in x 10.2 in x 1.8 in'
 '5 in x 6 in x 9 in' '7.87 in x 10.87 in x 0.79 in' '3 in x 7 in x 3 in'
 '4 in x 3 in x 8 in' '4 in x 9 in x 3 in' '32 inches'
 '22 in x 21 in x 23 in' '54 in x 40 in x 29 in' '7 in x 10.25 in x 3 in'
 '6.7 in x 15.8 in x 4.1 in' '4 in x 12 in x 0.25 in'
 '0.9 in x 8.4 in x 0.4 in' '4.5 in x 26 in x 1 in'
 '17.0 in x 15.0 in x 9.0 in' '7 in x 15 in x 11 in'
 '14 in x 19 in x 2 in' '4.09 in x 7.19 in x 6.53 in'
 '5 in x 13 in x 7 in' '6 in x 13 in x 5 in' '3 in x 5 in x 8 in'
 '7 in x 10 in x 2.1 in' '1 in x 1 in x 5 in' '7 in x 13.5 in x 5 in'
 '13.78 in x 9.45 in x 4.92 in' '12.8 in x 7.8 in x 4.92 in'
 '11 in x 12.5 in x 5.5 in' '10.75 in x 15 in x 2.5 in'

```
'4 in x 7 in x 2 in' '6 in x 9 in x 0.5 in' '5.9 in x 10.9 in x 4.5 in'
'14 in x 20 in x 4 in' '1.46 in x 13.01 in x 0.51 in'
'5.3 in x 13.7 in x 4.4 in' '5.5 in x 11.4 in x 1.3 in'
'4 in x 1 in x 10 in' '8.8 in x 13 in x 0.9 in'
'21.0 in x 5.0 in x 15.0 in' '6.4 in x 6.6 in x 4.8 in']
Uniques: 278
Nans: 16303
```

```
[46]: for d in unique_dims:
        if not pd.isnull(d) and len(d.split('x')) != 3:
            print(d)
```

32 inches

This means almost all dimensions can be split by x except for '32 inches' one.

```
[47]: def process_dimensions(x):
        if pd.isnull(x):
            return x

        # deal with edge case
        if x == '32 inches':
            return 32.0

        x = x.lower()
        dims = x.split('x')
        dims = [d.replace(' ', '').replace('in', '') for d in dims]
        dims = map(float, dims)
        return sum(dims)

df['dimension'] = df['dimension'].apply(process_dimensions)
df['dimension']
```

```
[47]: 0      NaN
      1      NaN
      2      NaN
      3      NaN
      4      NaN
      ...
19310   NaN
19311   NaN
19312   3.0
19313   3.0
19314   3.0
Name: dimension, Length: 19294, dtype: float64
```

5.13 Cleaning prices.size column

Drop this column, since too few values and mostly very different.

```
[48]: get_unique_nans('prices.size', df)
```

```
[nan 'S' 'S (5-6)' '7.5' '9 D' '11 D' '10 D' '13 D' '7.5 D' '12 D' '8 D'
'9.5 D' '10.5 D' '8.5 D' '7 D' '11.5 D' '14 D' '15 D' '10.5 E' '8.5 E'
'11 E' '10 E' '12 E' '8 E' '7 E' '9.5 E' '9 E' '7.5 E' '11.5 E' '12.5 D'
'10.5' '13.0' 'XL (44-46)' '7' 'S (8-9)' 'S (7-8)' 'Large' '9' '6' 'XSM'
'11.5 EE' '6.5 D' '11 EE' '8 EE' '6 D' '7.5 EE' '9 EE' '10 EE' '9.5 EE'
'7 EE' '10.5 EE' '8.5 EE' '12 EE' '1' '16 D' '17 D' '42 mm without crown'
'6 E' '5.5 E' 'One Size' '6.5 E' '5 D' '5.5 D']
```

Uniques: 63

Nans: 18645

```
[49]: df['prices.size'].value_counts()
```

```
[49]: 8 D          37
      11 D         37
      10 D         37
      12 D         37
      13 D         37
      ..
      5 D           1
      10 EE          1
      S (5-6)         1
      One Size         1
      7 EE            1
      Name: prices.size, Length: 62, dtype: int64
```

```
[50]: df = df.drop('prices.size', axis=1)
```

5.14 Cleaning prices.color column

Drop this too, for same reason.

```
[51]: get_unique_nans('prices.color', df)
```

```
[nan 'Camouflage' 'Java' 'Black' 'Dark Brown' 'Navy' 'Green/Newport Navy'
'Frame color: Tortoise, Lens color: Grey' 'Tan'
'Frame color: Blue, Lens color: Blue' 'Wheat' 'Realtree' 'Grey'
'Frame color: Black, Lens color: Grey' 'Green/Brown' 'Brown/Blk' 'Silver'
'Frame color: Blue, Lens color: Orange'
'Frame color: Grey, Lens color: Orange'
'Frame color: Blue, Lens color: Grey' 'Shiny Black / gray gradient'
'Shiny-Matte Brown / Photochromic Polarized PhD Drivers' 'Greige'
'Olive Camo/Black' 'Gunmetal / Blue Lens' 'Black / Gray Lens'
'Multi-Color' 'Blue' 'Brown' 'Burnished Dark Brown' 'Orange Plaid' 'Gold'
'Black / Grey' 'Black to Grey Havana / Gray Gradient']
```

```
'Fuzzy Black / Grey Polarized' 'Frame color: Grey, Lens color: Grey'
'Frame color: Silver, Lens color: Blue'
'Demi Gloss Sand Silver / Grey Gradient Dark Grey Polar Lens'
'BLACK GREY' 'Green' 'CAMO' 'New Snuff/Cream' 'Charcoal Grey'
'Newport Navy/Cream' 'Violet / dark gray shaded' 'BLACK/GREY'
'Frame color: Black, Lens color: Blue' 'Regent Grey' 'Newport Navy'
'Polo Black' 'Pure White' 'Rose Gold' 'Snuff' 'Orange' 'Turquoise'
'Wine/Olive']
```

Uniques: 56

Nans: 18573

```
[52]: df = df.drop('prices.color', axis=1)
```

5.15 Cleaning prices.returnPolicy column

Ready for categorizing.

```
[53]: get_unique_nans('prices.returnPolicy', df)
```

[nan

'FREE SHIPPING & FREE RETURNS Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$44.95. DETAILS'

'The full amount will be credited to the original tender (credit card, gift card, or online gift certificate) if returned within 60 days of the ship date on your invoice and accompanied by the original invoice. The current selling price will be credited if returned after 60 days and accompanied by the original invoice.'

'Always free shipping, fittings & returns'

"Returning your sunglasses is easy at Sunglasses Shop. We know that sometimes what you get isn't quite right or just not what you expected, so simply fill out the returns form and send us the sunglasses. We'll arrange your exchange, process your refund and if you can't decide on something different we can help you make the right choice. At Sunglasses Shop we love all things eyewear and want you to love what you've bought just as much. That's why we want you to be nothing less than entirely satisfied and because your style matters."

'FREE SHIPPING & FREE RETURNS Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$24.95. DETAILS'

'FREE SHIPPING & FREE RETURNS' '90-Day'

'\$5 SHIPPING. Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$24.95. DETAILS'

'30 Day Money Back Guarantee']

Uniques: 10

Nans: 18314

```
[54]: df['prices.returnPolicy'].value_counts()
```

```
[54]: The full amount will be credited to the original tender (credit card, gift card,
or online gift certificate) if returned within 60 days of the ship date on your
```

invoice and accompanied by the original invoice. The current selling price will be credited if returned after 60 days and accompanied by the original invoice.
674

Returning your sunglasses is easy at Sunglasses Shop. We know that sometimes what you get isn't quite right or just not what you expected, so simply fill out the returns form and send us the sunglasses. We'll arrange your exchange, process your refund and if you can't decide on something different we can help you make the right choice. At Sunglasses Shop we love all things eyewear and want you to love what you've bought just as much. That's why we want you to be nothing less than entirely satisfied and because your style matters. 105

FREE SHIPPING & FREE RETURNS Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$44.95. DETAILS

98

Always free shipping, fittings & returns

39

FREE SHIPPING & FREE RETURNS

25

90-Day

23

FREE SHIPPING & FREE RETURNS Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$24.95. DETAILS

7

30 Day Money Back Guarantee

7

\$5 SHIPPING. Join SHOES.com as an Elite member for 365 days of Next Day shipping for a low, one-time payment of \$24.95. DETAILS

2

Name: prices.returnPolicy, dtype: int64

5.16 Calculating Average Review Score

I decided to calculate average rating for reviews with rating. If there is no rating, return zero.

```
[55]: no_reviews_idx = df['reviews'].isnull()
```

```
[56]: df['reviews'][~no_reviews_idx]
```

```
[56]: 52      [{"date": "2015-09-18T00:00:00Z", "dateAdded": "2...
53      [{"date": "2015-09-18T00:00:00Z", "dateAdded": "2...
54      [{"date": "2015-09-18T00:00:00Z", "dateAdded": "2...
109     [{"date": "2015-10-16T00:00:00Z", "dateAdded": "2...
119     [{"date": "2015-12-16T00:00:00Z", "dateAdded": "2...
      ...
19278   [{"date": "2015-05-23T00:00:00Z", "dateAdded": "2...
19292   [{"date": "2015-11-30T00:00:00Z", "dateAdded": "2...
19293   [{"date": "2015-11-30T00:00:00Z", "dateAdded": "2...
19294   [{"date": "2015-11-30T00:00:00Z", "dateAdded": "2...
19298   [{"dateAdded": "2017-01-27T01:33:09Z", "dateSeen..."
```

Name: reviews, Length: 1659, dtype: object

```
[57]: def parse_review(x):
      if pd.isnull(x):
          return x

      rating = 0.0
      try:
          reviews = json.loads(x)
          for r in reviews:
              if 'rating' in r.keys():
                  rating += float(r['rating'])
          return rating / len(reviews)
      except:
          return rating

df['reviews'] = df['reviews'].apply(parse_review)
df['reviews'].value_counts()
```

```
[57]: 0.000000    813
      5.000000     94
      2.687500     66
      2.500000     44
      2.000000     43
      ...
      1.875000      1
      3.187500      1
      4.527778      1
      3.500000      1
      4.482759      1
      Name: reviews, Length: 70, dtype: int64
```

5.17 Examining and splitting the data

```
[58]: df
```

```
[58]:
```

	brand	colors	dimension	manufacturer	prices.condition \
0	Josmo	NaN	NaN	NaN	NaN
1	Josmo	NaN	NaN	NaN	new
2	SERVUS BY HONEYWELL	NaN	NaN	NaN	new
3	SERVUS BY HONEYWELL	NaN	NaN	NaN	new
4	SERVUS BY HONEYWELL	NaN	NaN	NaN	new
...
19310	Xray, XRay	black	NaN	NaN	New with box
19311	Fila	NaN	NaN	unknown	new
19312	Seek Optics	NaN	3.0	NaN	NaN
19313	Seek Optics	NaN	3.0	NaN	NaN
19314	Seek Optics	NaN	3.0	NaN	NaN

	prices.currency	prices.isSale	prices.merchant	\
0	USD	True	NaN	
1	USD	False	UnbeatableSale - Walmart.com	
2	USD	False	SIM Supply Inc - Walmart.com	
3	USD	False	SIM Supply Inc - Walmart.com	
4	USD	False	SIM Supply Inc - Walmart.com	
...	
19310	USD	False	NaN	
19311	USD	False	ULTIMATE SHOE ZONE - Walmart.com	
19312	USD	True	Seek Optics, Inc	
19313	USD	True	Seek Optics, Inc	
19314	USD	False	Seek Optics, Inc	

	prices.offer	prices.returnPolicy	prices.shipping	reviews	sizes	\
0	REDUCED USD 12.10	NaN	NaN	NaN	NaN	
1	REDUCED USD 12.10	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	
...	
19310	NaN	NaN	not free	NaN	1.0	
19311	NaN	NaN	NaN	NaN	NaN	
19312	NaN	NaN	NaN	NaN	NaN	
19313	NaN	NaN	NaN	NaN	NaN	
19314	NaN	NaN	NaN	NaN	NaN	

	price
0	39.89
1	51.99
2	40.02
3	50.31
4	46.26
...	...
19310	29.95
19311	64.95
19312	32.48
19313	31.48
19314	72.00

[19294 rows x 14 columns]

5.18 Get rid of price outliers

```
[59]: df = df.drop(df[df.price > 2000].index)
df
```

```
[59]:
```

	brand	colors	dimension	manufacturer	prices.condition	\
0	Josmo	NaN	NaN	NaN	NaN	
1	Josmo	NaN	NaN	NaN	new	
2	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
3	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
4	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
...	
19310	Xray, XRay	black	NaN	NaN	New with box	
19311	Fila	NaN	NaN	unknown	new	
19312	Seek Optics	NaN	3.0	NaN	NaN	
19313	Seek Optics	NaN	3.0	NaN	NaN	
19314	Seek Optics	NaN	3.0	NaN	NaN	

	prices.currency	prices.isSale	prices.merchant	\
0	USD	True	NaN	
1	USD	False	UnbeatableSale - Walmart.com	
2	USD	False	SIM Supply Inc - Walmart.com	
3	USD	False	SIM Supply Inc - Walmart.com	
4	USD	False	SIM Supply Inc - Walmart.com	
...	
19310	USD	False	NaN	
19311	USD	False	ULTIMATE SHOE ZONE - Walmart.com	
19312	USD	True	Seek Optics, Inc	
19313	USD	True	Seek Optics, Inc	
19314	USD	False	Seek Optics, Inc	

	prices.offer	prices.returnPolicy	prices.shipping	reviews	sizes	\
0	REDUCED USD 12.10	NaN	NaN	NaN	NaN	
1	REDUCED USD 12.10	NaN	NaN	NaN	NaN	
2	NaN	NaN	NaN	NaN	NaN	
3	NaN	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	NaN	
...	
19310	NaN	NaN	not free	NaN	1.0	
19311	NaN	NaN	NaN	NaN	NaN	
19312	NaN	NaN	NaN	NaN	NaN	
19313	NaN	NaN	NaN	NaN	NaN	
19314	NaN	NaN	NaN	NaN	NaN	

	price
0	39.89
1	51.99
2	40.02

```

3      50.31
4      46.26
...      ...
19310   29.95
19311   64.95
19312   32.48
19313   31.48
19314   72.00

```

[19270 rows x 14 columns]

```
[60]: X, y = df.drop('price', axis=1), df['price']
```

```
[61]: X
```

```
[61]:
```

	brand	colors	dimension	manufacturer	prices.condition	\
0	Josmo	NaN	NaN	NaN	NaN	
1	Josmo	NaN	NaN	NaN	new	
2	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
3	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
4	SERVUS BY HONEYWELL	NaN	NaN	NaN	new	
...	
19310	Xray, XRay	black	NaN	NaN	New with box	
19311	Fila	NaN	NaN	unknown	new	
19312	Seek Optics	NaN	3.0	NaN	NaN	
19313	Seek Optics	NaN	3.0	NaN	NaN	
19314	Seek Optics	NaN	3.0	NaN	NaN	

	prices.currency	prices.isSale	prices.merchant	\
0	USD	True	NaN	
1	USD	False	UnbeatableSale - Walmart.com	
2	USD	False	SIM Supply Inc - Walmart.com	
3	USD	False	SIM Supply Inc - Walmart.com	
4	USD	False	SIM Supply Inc - Walmart.com	
...	
19310	USD	False	NaN	
19311	USD	False	ULTIMATE SHOE ZONE - Walmart.com	
19312	USD	True	Seek Optics, Inc	
19313	USD	True	Seek Optics, Inc	
19314	USD	False	Seek Optics, Inc	

	prices.offer	prices.returnPolicy	prices.shipping	reviews	sizes
0	REDUCED USD 12.10	NaN	NaN	NaN	NaN
1	REDUCED USD 12.10	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	NaN
...

19310	NaN	NaN	not free	NaN	1.0
19311	NaN	NaN	NaN	NaN	NaN
19312	NaN	NaN	NaN	NaN	NaN
19313	NaN	NaN	NaN	NaN	NaN
19314	NaN	NaN	NaN	NaN	NaN

[19270 rows x 13 columns]

[62]:

y

[62]:

0	39.89
1	51.99
2	40.02
3	50.31
4	46.26
...	
19310	29.95
19311	64.95
19312	32.48
19313	31.48
19314	72.00

Name: price, Length: 19270, dtype: float64

[63]:

X.describe()

[63]:

	dimension	reviews	sizes
count	2989.000000	1659.000000	5966.000000
mean	14.690995	1.458701	4.260979
std	13.227951	1.737385	5.865847
min	0.030000	0.000000	1.000000
25%	3.000000	0.000000	1.000000
50%	13.000000	0.888889	2.000000
75%	20.000000	2.687500	5.000000
max	123.000000	5.000000	60.000000

[64]:

X.columns

[64]:

Index(['brand', 'colors', 'dimension', 'manufacturer', 'prices.condition', 'prices.currency', 'prices.isSale', 'prices.merchant', 'prices.offer', 'prices.returnPolicy', 'prices.shipping', 'reviews', 'sizes'], dtype='object')

[65]:

```
categorical=[
    'brand',
    'colors',
    'manufacturer',
    'prices.condition',
    'prices.currency',
    'prices.isSale',
    'prices.merchant',
```

```

    'prices.offer',
    'prices.returnPolicy',
    'prices.shipping',
]

```

```
[ ]:
```

```

[66]: import numpy as np
import category_encoders as ce
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error

```

```

[67]: # apply random permutation to the data (required by CatBoost)
perm = np.random.permutation(len(X))
X = X.iloc[perm].reset_index(drop=True)
y = y.iloc[perm].reset_index(drop=True)

```

```

[68]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
↳shuffle=False)

```

```

[69]: ce_cat = ce.CatBoostEncoder(cols=categorical)

```

```

[70]: X_train

```

```

[70]:          brand          colors \
0          PUMA          NaN
1          Nike          NaN
2  Concepts Sports          NaN
3          RBX  dunkball,bklynball,gateball,lightningball,rimb...
4      Spy Optic          black
...          ...          ...
15411      sumaclife  gray,red,pink,orange,green,blue
15412          Laredo          brown
15413          NaN          green
15414          14k co.          NaN
15415  Mercanti Fiorentini          brown

```

```

          dimension  manufacturer  prices.condition  prices.currency \
0          NaN          Puma          NaN          USD
1          NaN          NaN  New with box          USD
2          12.0          NaN          NaN          USD
3          NaN          NaN          NaN          USD
4          NaN          NaN  New with tags          AUD
...          ...          ...          ...          ...
15411          NaN          NaN          NaN          USD
15412          NaN          NaN          new          USD
15413          NaN          Dunlop          new          EUR
15414          NaN          NaN          NaN          USD
15415          NaN  Mercanti Fiorentini          new          USD

```

	prices.isSale	prices.merchant	prices.offer \
0	True	NaN	REDUCED USD 18.01
1	False	graduate08in82 - Ebay.com	NaN
2	False	oh!Dreamy	NaN
3	True	Wireless4Wireless	NaN
4	False	NaN	NaN
...
15411	True	Sears.com	NaN
15412	False	ApparelSave - Walmart.com	NaN
15413	False	Universal Textiles & Footwear	23% off
15414	False	Sears.com	NaN
15415	False	ApparelSave - Walmart.com	REDUCED USD 133.01

	prices.returnPolicy	prices.shipping	reviews	sizes
0	NaN	NaN	NaN	NaN
1	NaN	not free	NaN	1.0
2	NaN	NaN	NaN	NaN
3	NaN	free	2.444444	NaN
4	NaN	not free	NaN	NaN
...
15411	NaN	NaN	NaN	NaN
15412	NaN	NaN	NaN	12.0
15413	NaN	not free	NaN	NaN
15414	NaN	NaN	NaN	1.0
15415	NaN	NaN	NaN	1.0

[15416 rows x 13 columns]

```
[71]: X_train = ce_cat.fit_transform(X_train, y_train)
X_train
```

```
[71]:
```

	brand	colors	dimension	manufacturer	prices.condition \
0	108.033927	108.033927	-999.9	108.033927	108.033927
1	108.033927	90.011963	-999.9	108.033927	108.033927
2	108.033927	83.337976	12.0	89.011963	90.011963
3	108.033927	108.033927	-999.9	67.671309	68.337976
4	108.033927	108.033927	-999.9	53.000982	108.033927
...
15411	46.740982	69.011963	-999.9	118.126095	106.750840
15412	111.156305	93.561804	-999.9	118.116300	86.898632
15413	82.693912	82.218471	-999.9	36.236321	86.905012
15414	167.629426	99.439379	-999.9	118.117276	106.734435
15415	160.571990	93.616037	-999.9	160.571990	86.894162

	prices.currency	prices.isSale	prices.merchant	prices.offer \
0	108.033927	108.033927	108.033927	108.033927
1	90.011963	108.033927	108.033927	108.033927
2	83.337976	89.011963	108.033927	89.011963

3	68.750982	90.011963	108.033927	67.671309
4	108.033927	67.671309	90.011963	53.000982
...
15411	109.188109	77.595565	114.019879	124.763575
15412	109.182005	120.533609	90.588088	124.753566
15413	104.472472	120.534293	29.838436	34.626488
15414	109.183278	120.524817	113.886541	124.753873
15415	109.184656	120.525644	90.611558	188.341309

	prices.returnPolicy	prices.shipping	reviews	sizes
0	108.033927	108.033927	-999.900000	-999.9
1	90.011963	108.033927	-999.900000	1.0
2	83.337976	90.011963	-999.900000	-999.9
3	68.750982	108.033927	2.444444	-999.9
4	56.798785	89.011963	-999.900000	-999.9
...
15411	92.544690	90.097270	-999.900000	-999.9
15412	92.539659	90.090673	-999.900000	12.0
15413	92.542083	81.179863	-999.900000	-999.9
15414	92.536920	90.094188	-999.900000	1.0
15415	92.539450	90.097847	-999.900000	1.0

[15416 rows x 13 columns]

```
[72]: def logify(data):
      return np.log(1 + data)

      def unlogify(data):
          return np.exp(data) - 1
```

```
[73]: y_train = logify(y_train)
```

```
[74]: from sklearn.ensemble import GradientBoostingRegressor
```

```
[75]: model = GradientBoostingRegressor(max_leaf_nodes=None, n_estimators=50)
```

```
[76]: model.fit(X_train, y_train)
```

```
[76]: GradientBoostingRegressor(alpha=0.9, criterion='friedman_mse', init=None,
                                learning_rate=0.1, loss='ls', max_depth=3,
                                max_features=None, max_leaf_nodes=None,
                                min_impurity_decrease=0.0, min_impurity_split=None,
                                min_samples_leaf=1, min_samples_split=2,
                                min_weight_fraction_leaf=0.0, n_estimators=50,
                                n_iter_no_change=None, presort='auto',
                                random_state=None, subsample=1.0, tol=0.0001,
                                validation_fraction=0.1, verbose=0, warm_start=False)
```

```
[77]: X_test
```

[77]:

	brand \
15416	PUMA
15417	Mephisto
15418	BEKINA
15419	Nike
15420	Tommie Copper
...	...
19265	HOOey
19266	PUMA
19267	Silhouette
19268	DVS
19269	Denim & Supply Ralph Lauren

	colors	dimension \
15416	black	NaN
15417	blackpalace,darkbrownpalace	NaN
15418	green	NaN
15419	blue	NaN
15420	yellow	NaN
...
19265	NaN	NaN
19266	black	NaN
19267	blackcarbonfiber	NaN
19268	blacknubuckdeegan,blackcanvas,black,blackwhite...	NaN
19269	blacks	NaN

	manufacturer	prices.condition	prices.currency	prices.isSale \
15416	Puma	new	USD	False
15417	NaN	NaN	USD	True
15418	BEKINA	new	USD	False
15419	NaN	New with box	USD	False
15420	NaN	New with tags	USD	False
...
19265	NaN	NaN	USD	True
19266	Puma	new	USD	True
19267	NaN	New with tags	USD	False
19268	NaN	NaN	USD	False
19269	NaN	New with tags	USD	False

	prices.merchant	prices.offer	prices.returnPolicy \
15416	ApparelSave - Walmart.com	REDUCED USD 43.02	NaN
15417	NaN	NaN	NaN
15418	Zoro - Walmart.com	NaN	NaN
15419	NaN	NaN	NaN
15420	NaN	winning bid	NaN
...
19265	Sears.com	NaN	NaN

19266	ApparelSave - Walmart.com	REDUCED	USD	15.01	NaN
19267		NaN		NaN	NaN
19268	Amazon.com			NaN	NaN
19269		NaN	winning bid		NaN

	prices.shipping	reviews	sizes
15416	NaN	NaN	12.0
15417	NaN	1.888889	NaN
15418	NaN	NaN	1.0
15419	not free	NaN	1.0
15420	not free	NaN	NaN
...
19265	NaN	NaN	13.0
19266	NaN	NaN	1.0
19267	free	NaN	NaN
19268	NaN	1.233333	NaN
19269	free	NaN	NaN

[3854 rows x 13 columns]

```
[78]: X_test_transformed = ce_cat.transform(X_test)
X_test_transformed
```

```
[78]:
```

	brand	colors	dimension	manufacturer	prices.condition	\
15416	70.627786	139.809532	NaN	69.349229	86.926456	
15417	241.063760	108.033927	NaN	118.118406	106.738699	
15418	163.831309	81.636119	NaN	163.831309	86.926456	
15419	100.376315	73.363645	NaN	118.118406	98.301166	
15420	10.818928	52.896261	NaN	118.118406	70.106751	
...	
19265	236.958214	99.443924	NaN	118.118406	106.738699	
19266	70.627786	139.809532	NaN	69.349229	86.926456	
19267	108.033927	108.033927	NaN	118.118406	70.106751	
19268	74.314881	74.314881	NaN	118.118406	106.738699	
19269	108.033927	47.149358	NaN	118.118406	70.106751	

	prices.currency	prices.isSale	prices.merchant	prices.offer	\
15416	109.197224	120.541608	90.739702	94.116741	
15417	109.197224	77.582501	82.222361	124.754328	
15418	109.197224	120.541608	101.879040	124.754328	
15419	109.197224	120.541608	82.222361	124.754328	
15420	109.197224	120.541608	82.222361	62.278692	
...	
19265	109.197224	77.582501	113.908493	124.754328	
19266	109.197224	77.582501	90.739702	61.409665	
19267	109.197224	120.541608	82.222361	124.754328	
19268	109.197224	120.541608	165.323759	124.754328	
19269	109.197224	120.541608	82.222361	62.278692	

	prices.returnPolicy	prices.shipping	reviews	sizes
15416	92.553286	90.116840	NaN	12.0
15417	92.553286	90.116840	1.888889	NaN
15418	92.553286	90.116840	NaN	1.0
15419	92.553286	81.132529	NaN	1.0
15420	92.553286	81.132529	NaN	NaN
...
19265	92.553286	90.116840	NaN	13.0
19266	92.553286	90.116840	NaN	1.0
19267	92.553286	178.192042	NaN	NaN
19268	92.553286	90.116840	1.233333	NaN
19269	92.553286	178.192042	NaN	NaN

[3854 rows x 13 columns]

```
[79]: X.describe().columns
```

```
[79]: Index(['dimension', 'reviews', 'sizes'], dtype='object')
```

```
[80]: cols_to_fill_nas = ['dimension', 'reviews', 'sizes']
for c in cols_to_fill_nas:
    X_test_transformed[c] = X_test_transformed[c].fillna(-999.9)
X_test_transformed
```

```
[80]:
```

	brand	colors	dimension	manufacturer	prices.condition \
15416	70.627786	139.809532	-999.9	69.349229	86.926456
15417	241.063760	108.033927	-999.9	118.118406	106.738699
15418	163.831309	81.636119	-999.9	163.831309	86.926456
15419	100.376315	73.363645	-999.9	118.118406	98.301166
15420	10.818928	52.896261	-999.9	118.118406	70.106751
...
19265	236.958214	99.443924	-999.9	118.118406	106.738699
19266	70.627786	139.809532	-999.9	69.349229	86.926456
19267	108.033927	108.033927	-999.9	118.118406	70.106751
19268	74.314881	74.314881	-999.9	118.118406	106.738699
19269	108.033927	47.149358	-999.9	118.118406	70.106751

	prices.currency	prices.isSale	prices.merchant	prices.offer \
15416	109.197224	120.541608	90.739702	94.116741
15417	109.197224	77.582501	82.222361	124.754328
15418	109.197224	120.541608	101.879040	124.754328
15419	109.197224	120.541608	82.222361	124.754328
15420	109.197224	120.541608	82.222361	62.278692
...
19265	109.197224	77.582501	113.908493	124.754328
19266	109.197224	77.582501	90.739702	61.409665
19267	109.197224	120.541608	82.222361	124.754328
19268	109.197224	120.541608	165.323759	124.754328

19269	109.197224	120.541608	82.222361	62.278692
-------	------------	------------	-----------	-----------

	prices.returnPolicy	prices.shipping	reviews	sizes
15416	92.553286	90.116840	-999.900000	12.0
15417	92.553286	90.116840	1.888889	-999.9
15418	92.553286	90.116840	-999.900000	1.0
15419	92.553286	81.132529	-999.900000	1.0
15420	92.553286	81.132529	-999.900000	-999.9
...
19265	92.553286	90.116840	-999.900000	13.0
19266	92.553286	90.116840	-999.900000	1.0
19267	92.553286	178.192042	-999.900000	-999.9
19268	92.553286	90.116840	1.233333	-999.9
19269	92.553286	178.192042	-999.900000	-999.9

[3854 rows x 13 columns]

```
[81]: preds = model.predict(X_test_transformed)
```

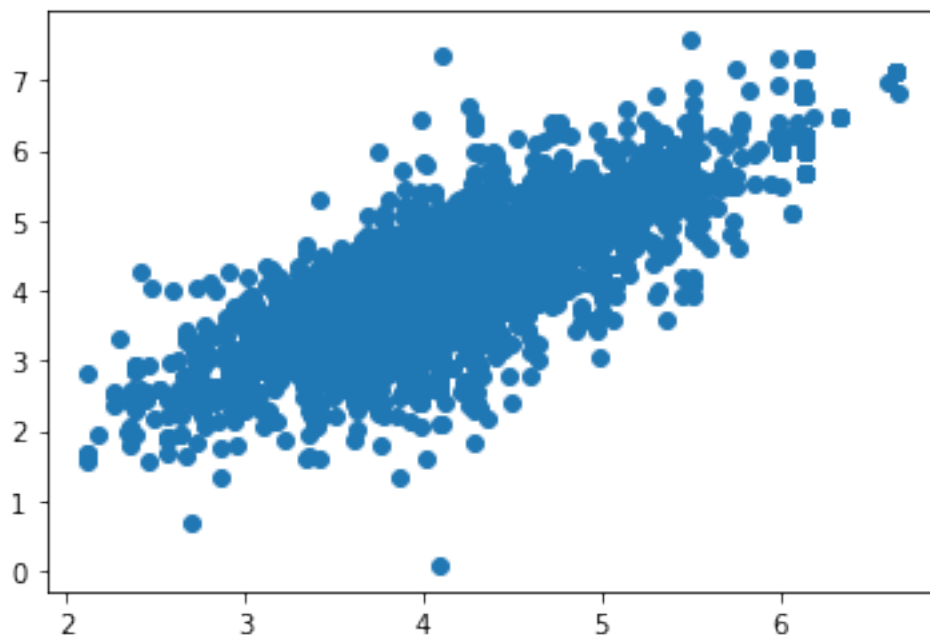
```
[82]: mean_squared_error(preds, logify(y_test))
```

```
[82]: 0.32169262519768926
```

```
[83]: import matplotlib.pyplot as plt
      %matplotlib inline
```

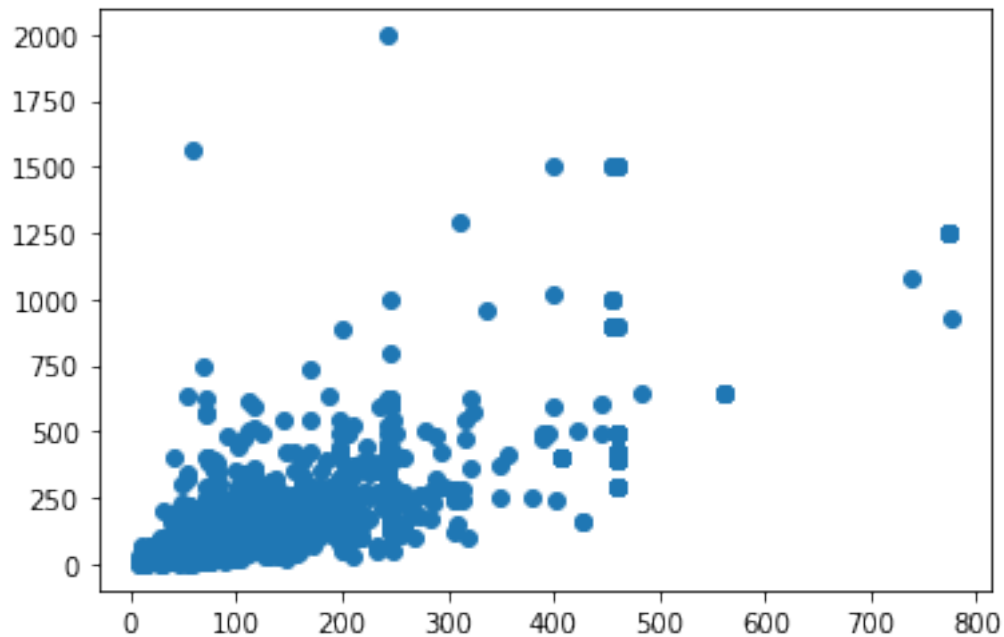
```
[84]: plt.scatter(preds, logify(y_test))
```

```
[84]: <matplotlib.collections.PathCollection at 0x1a20f87950>
```



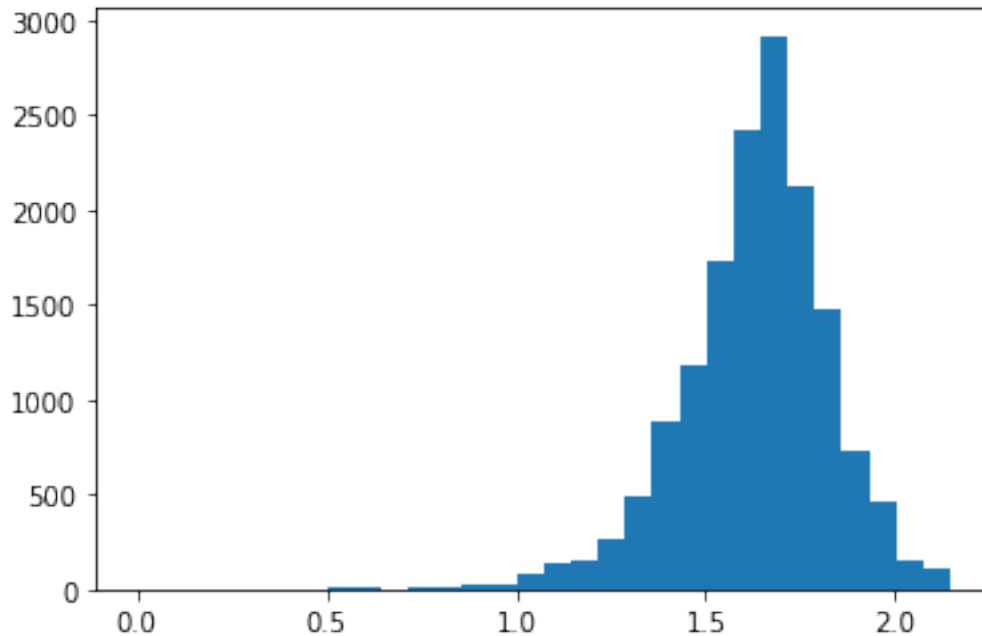
```
[85]: plt.scatter(unlogify(preds), y_test)
```

```
[85]: <matplotlib.collections.PathCollection at 0x1a2113b790>
```



```
[86]: plt.hist(np.log(1+y_train), bins=30)
```

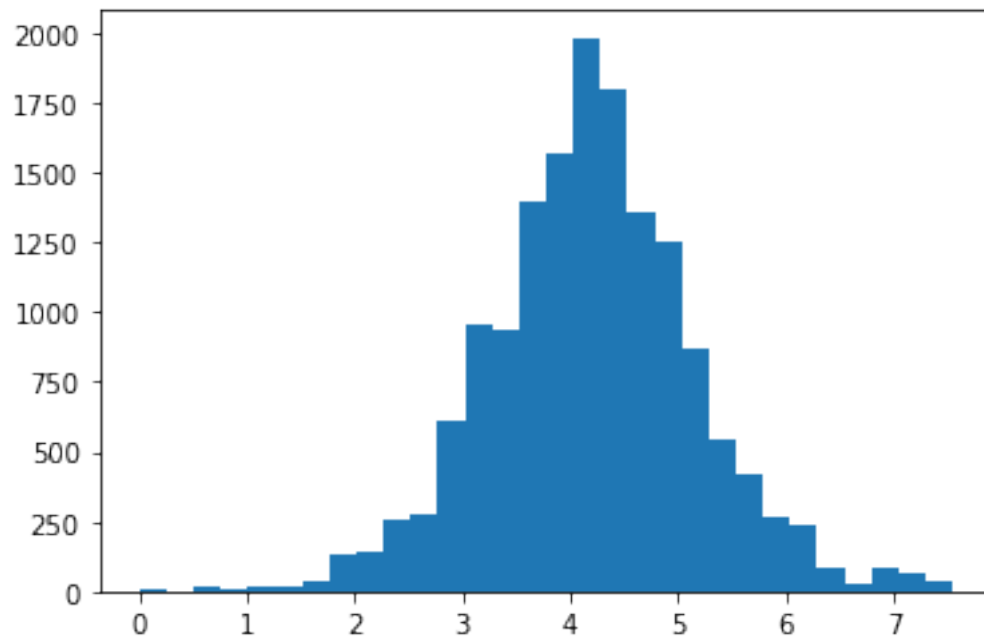
```
[86]: (array([4.000e+00, 4.000e+00, 0.000e+00, 0.000e+00, 0.000e+00, 0.000e+00,
        0.000e+00, 1.400e+01, 6.000e+00, 1.000e+00, 1.100e+01, 1.000e+01,
        2.000e+01, 3.200e+01, 7.500e+01, 1.360e+02, 1.590e+02, 2.670e+02,
        4.930e+02, 8.840e+02, 1.185e+03, 1.723e+03, 2.422e+03, 2.917e+03,
        2.128e+03, 1.475e+03, 7.340e+02, 4.550e+02, 1.500e+02, 1.110e+02]),
array([0.          , 0.07153157, 0.14306314, 0.21459471, 0.28612628,
        0.35765785, 0.42918942, 0.50072099, 0.57225256, 0.64378413,
        0.7153157 , 0.78684727, 0.85837884, 0.92991042, 1.00144199,
        1.07297356, 1.14450513, 1.2160367 , 1.28756827, 1.35909984,
        1.43063141, 1.50216298, 1.57369455, 1.64522612, 1.71675769,
        1.78828926, 1.85982083, 1.9313524 , 2.00288397, 2.07441554,
        2.14594711]),
<a list of 30 Patch objects>)
```



```
[87]: lg = np.log(1+y_train)
      un = np.exp(lg) - 1
```

```
[88]: plt.hist(un, bins=30)
```

```
[88]: (array([ 8.,  0., 14.,  7., 15., 22., 39., 134., 141.,
            255., 279., 614., 957., 935., 1395., 1573., 1983., 1803.,
            1356., 1249., 870., 547., 423., 268., 235., 80., 28.,
            81., 67., 38.]),
      array([0., 0.25167118, 0.50334236, 0.75501353, 1.00668471,
            1.25835589, 1.51002707, 1.76169825, 2.01336942, 2.2650406 ,
            2.51671178, 2.76838296, 3.02005414, 3.27172532, 3.52339649,
            3.77506767, 4.02673885, 4.27841003, 4.53008121, 4.78175238,
            5.03342356, 5.28509474, 5.53676592, 5.7884371 , 6.04010827,
            6.29177945, 6.54345063, 6.79512181, 7.04679299, 7.29846416,
            7.55013534]),
      <a list of 30 Patch objects>)
```



[: