

Amazon Apparel Recommendations



▼ [4.3] Overview of the data

```
#import all the necessary packages.
```

```
from PIL import Image
# import PIL.Image
import requests
from io import BytesIO
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
import pandas as pd
import warnings
from bs4 import BeautifulSoup
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk
import math
import time
import re
import os
import seaborn as sns
from collections import Counter
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.metrics import pairwise_distances
from matplotlib import gridspec
from scipy.sparse import hstack
import plotly
import plotly.figure_factory as ff
from plotly.graph_objs import Scatter, Layout
```

plotly.offline.init_notebook_mode(connected=True)
warnings.filterwarnings("ignore")



```
# we have give a json file which consists of all information about  
# the products  
# loading the data using pandas' read_json file.  
data = pd.read_json('tops_fashion.json')
```

```
print ('Number of data points : ', data.shape[0], \  
      'Number of features/variables:', data.shape[1])
```



Number of data points : 183138 Number of features/variables: 19

▼ Terminology:

What is a dataset?

Rows and columns

Data-point

Feature/variable

```
# each product/item has 19 features in the raw dataset.
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
'editorial_review', 'editorial_review', 'formatted_price',  
'large_image_url', 'manufacturer', 'medium_image_url', 'model',  
'product_type_name', 'publisher', 'reviews', 'sku', 'small_image_url',  
'title'],  
dtype='object')
```

Of these 19 features, we will be using only 6 features in this workshop. 1. asin (Amazon standard id which the product belongs to) 3. color (Color information of apparel, it can contain many colors as a product_type_name (type of the apparel, ex: SHIRT/TSHIRT) 5. medium_image_url (url of the image) formatted_price (price of the product)

```
data = data[['asin', 'brand', 'color', 'medium_image_url', 'product_type_name', 'title', 'for  
print ('Number of data points : ', data.shape[0], \  
      'Number of features:', data.shape[1])  
data.head() # prints the top rows in the table.
```



Number of data points : 183138 Number of features: 7

	asin	brand	color	medium_image_url	product_type_name
0	B016I2TS4W	FNC7C	None	https://images-na.ssl-images-amazon.com/images...	Shirt
1	B01N49AI08	FIG Clothing	None	https://images-na.ssl-images-amazon.com/images...	Shirt
2	B01JDPCOHO	FIG Clothing	None	https://images-na.ssl-images-amazon.com/images...	Shirt
3	B01N19U5H5	Focal18	None	https://images-na.ssl-images-amazon.com/images...	Shirt
4	B004GSI2OS	FeatherLite	Onyx Black/Stone	https://images-na.ssl-images-amazon.com/images...	Shirt

- ▼ [5.1] Missing data for various features.
- ▼ Basic stats for the feature: product_type_name

```
# We have total 72 unique type of product_type_names
print(data['product_type_name'].describe())
```

```
# 91.62% (167794/183138) of the products are shirts,
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

```
unique      72
top        SHIRT
freq     167794
Name: product_type_name, dtype: object
```

```
# names of different product types
print(data['product_type_name'].unique())
```



```
['SHIRT' 'SWEATER' 'APPAREL' 'OUTDOOR_RECREATION_PRODUCT'
 'BOOKS_1973_AND_LATER' 'PANTS' 'HAT' 'SPORTING_GOODS' 'DRESS' 'UNDERWEAR'
 'SKIRT' 'OUTERWEAR' 'BRA' 'ACCESSORY' 'ART_SUPPLIES' 'SLEEPWEAR'
 'ORCA_SHIRT' 'HANDBAG' 'PET_SUPPLIES' 'SHOES' 'KITCHEN' 'ADULT_COSTUME'
 'HOME_BED_AND_BATH' 'MISC_OTHER' 'BLAZER' 'HEALTH_PERSONAL_CARE'
 'TOYS_AND_GAMES' 'SWIMWEAR' 'CONSUMER_ELECTRONICS' 'SHORTS' 'HOME'
 'AUTO_PART' 'OFFICE_PRODUCTS' 'ETHNIC_WEAR' 'BEAUTY'
 'INSTRUMENT_PARTS_AND_ACCESSORIES' 'POWERSPORTS_PROTECTIVE_GEAR' 'SHIRTS'
 'ABIS_APPAREL' 'AUTO_ACCESSORY' 'NONAPPARELMISC' 'TOOLS' 'BABY_PRODUCT'
 'SOCKSHOSIERY' 'POWERSPORTS RIDING SHIRT' 'EYEWEAR' 'SUIT'
 'OUTDOOR_LIVING' 'POWERSPORTS RIDING JACKET' 'HARDWARE' 'SAFETY_SUPPLY'
 'ABIS_DVD' 'VIDEO_DVD' 'GOLF_CLUB' 'MUSIC_POPULAR_VINYL'
 'HOME_FURNITURE_AND_DECOR' 'TABLET_COMPUTER' 'GUILD_ACCESSORIES'
 'ABIS_SPORTS' 'ART_AND_CRAFT_SUPPLY' 'BAG' 'MECHANICAL_COMPONENTS'
 'SOUND_AND_RECORDING_EQUIPMENT' 'COMPUTER_COMPONENT' 'JEWELRY'
 'BUILDING_MATERIAL' 'LUGGAGE' 'BABY_COSTUME' 'POWERSPORTS_VEHICLE_PART'
 'PROFESSIONAL_HEALTHCARE' 'SEEDS_AND_PLANTS' 'WIRELESS_ACCESSORY']
```

```
# find the 10 most frequent product_type_names.
product_type_count = Counter(list(data['product_type_name']))
product_type_count.most_common(10)
```

👤 [('SHIRT', 167794),
 ('APPAREL', 3549),
 ('BOOKS_1973_AND_LATER', 3336),
 ('DRESS', 1584),
 ('SPORTING_GOODS', 1281),
 ('SWEATER', 837),
 ('OUTERWEAR', 796),
 ('OUTDOOR RECREATION PRODUCT', 729)]

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

▼ Basic stats for the feature: brand

```
# there are 10577 unique brands
print(data['brand'].describe())

# 183138 - 182987 = 151 missing values.
```

👤 count 182987
 unique 10577
 top Zago
 freq 223
 Name: brand, dtype: object

```
brand_count = Counter(list(data['brand']))
brand_count.most_common(10)
```

👤

```
[('Zago', 223),  
 ('XQS', 222),  
 ('Yayun', 215),  
 ('YUNY', 198),  
 ('XiaoTianXin-women clothes', 193),  
 ('Generic', 192),  
 ('Boohoo', 190),  
 ('Alion', 188),  
 ('Abetteric', 187),  
 ('TheMogan', 187)]
```

▼ Basic stats for the feature: color

```
print(data['color'].describe())
```

```
# we have 7380 unique colors  
# 7.2% of products are black in color  
# 64956 of 183138 products have brand information. That's approx 35.4%.
```

```
👤 count      64956  
unique      7380  
top        Black  
freq       13207  
Name: color, dtype: object
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
👤 [(None, 118182),  
 ('Black', 13207),  
 ('White', 8616),  
 ('Blue', 3570),  
 ('Red', 2289),  
 ('Pink', 1842),  
 ('Grey', 1499),  
 ('*', 1388),  
 ('Green', 1258),  
 ('Multi', 1203)]
```

▼ Basic stats for the feature: formatted_price

```
print(data['formatted_price'].describe())
```

```
# Only 28,395 (15.5% of whole data) products with price information
```

```
👤
```

```
count      28395
unique     3135
top       $19.99
freq       945
Name: formatted_price, dtype: object
```

```
price_count = Counter(list(data['formatted_price']))
price_count.most_common(10)
```

👤 [(None, 154743),
 ('\$19.99', 945),
 ('\$9.99', 749),
 ('\$9.50', 601),
 ('\$14.99', 472),
 ('\$7.50', 463),
 ('\$24.99', 414),
 ('\$29.99', 370),
 ('\$8.99', 343),
 ('\$9.01', 336)]

▼ Basic stats for the feature: title

```
print(data['title'].describe())
```

👤 count 183138
unique 175985
top Nakoda Cotton Self Print Straight Kurti For Women
freq 77

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

▼ 6. Text pre-processing

```
data = pd.read_pickle('pickels/16k_appral_data')
```

```
# we use the list of stop words that are downloaded from nltk lib.
stop_words = set(stopwords.words('english'))
print ('list of stop words:', stop_words)
```

```
def nlp_preprocessing(total_text, index, column):
    if type(total_text) is not int:
        string = ""
        for words in total_text.split():
            # remove the special chars in review like '#$@!%^&*()_+-~?>< etc.
            word = ("".join(e for e in words if e.isalnum()))
            # Convert all letters to lower-case
            word = word.lower()
            # stop-word removal
            if word not in stop_words:
                string += word + " "
```

```
    if not word in stop_words:
        string += word + " "
    data[column][index] = string
```

list of stop words: {"you'll", 'herself', 'should', 'once', 'of', 'doing', 'down', 'were'}

```
start_time = time.clock()
# we take each title and we text-preprocess it.
for index, row in data.iterrows():
    nlp_preprocessing(row['title'], index, 'title')
# we print the time it took to preprocess whole titles
print(time.clock() - start_time, "seconds")
```

5.402977 seconds

data.shape

(16042, 7)

data.head()

	asin	brand	color	medium_image_url
4	B004GSI2OS	FeatherLite	Onyx Black/Stone	https://images-na.ssl-images-amazon.com/images/...
6	B012YX2ZPI	HX-Kingdom Fashion T-shirts	White	https://images-na.ssl-images-amazon.com/images/...
27	B014ICEJ1Q	FNC7C	Purple	https://images-na.ssl-images-amazon.com/images/...
46	B01NACPBG2	Fifth Degree	Black	https://images-na.ssl-images-amazon.com/images/...

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

data.to_pickle('pickels/16k_appral_data_preprocessed')

▼ Stemming

```
from nltk.stem.porter import *
stemmer = PorterStemmer()
print(stemmer.stem('arguing'))
print(stemmer.stem('fishing'))
```

We tried using stemming on our titles and it didnot work very well.

 argu
fish

▼ [8] Text based product similarity

```
data = pd.read_pickle('pickels/16k_apperal_data_preprocessed')
data.head()
```

	asin	brand	color	medium_image_url
4	B004GSI2OS	FeatherLite	Onyx Black/Stone	https://images-na.ssl-images-amazon.com/images...
6	B012YX2ZPI	HX-Kingdom Fashion T-shirts	White	https://images-na.ssl-images-amazon.com/images...
15	B003BSRPB0	FeatherLite	White	https://images-na.ssl-images-amazon.com/images...
27	B014ICEJ1Q	FNC7C	Purple	https://images-na.ssl-images-amazon.com/images...
46	B01NACPBG2	Fifth Degree	Black	https://images-na.ssl-images-amazon.com/images...

```
# Utility Functions which we will use through the rest of the workshop.
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

#Display an image

```
def display_img(url,ax,fig):
    # we get the url of the apparel and download it
    response = requests.get(url)
    img = Image.open(BytesIO(response.content))
    # we will display it in notebook
    plt.imshow(img)
```

#plotting code to understand the algorithm's decision.

```
def plot_heatmap(keys, values, labels, url, text):
    # keys: list of words of recommended title
    # values: len(values) == len(keys), values(i) represents the occurrence of the word k
    # labels: len(labels) == len(keys), the values of labels depends on the model we are
            # if model == 'bag of words': labels(i) = values(i)
            # if model == 'tfidf weighted bag of words': labels(i) = tfidf(keys(i))
            # if model == 'idf weighted bag of words': labels(i) = idf(keys(i))
    # url : apparel's url

    # we will devide the whole figure into two parts
    gs = gridspec.GridSpec(2, 2, width_ratios=[4,1], height_ratios=[4,1])
    fig = plt.figure(figsize=(12, 8))
```

```

fig = plt.figure(figsize=(12, 12))

# 1st, plotting heat map that represents the count of commonly occurred words in title2
ax = plt.subplot(gs[0])
# it displays a cell in white color if the word is intersection(lis of words of title
ax = sns.heatmap(np.array([values]), annot=np.array([labels]))
ax.set_xticklabels(keys) # set that axis labels as the words of title
ax.set_title(text) # apparel title

# 2nd, plotting image of the the apparel
ax = plt.subplot(gs[1])
# we don't want any grid lines for image and no labels on x-axis and y-axis
ax.grid(False)
ax.set_xticks([])
ax.set_yticks([])

# we call dispaly_img based with paramete url
display_img(url, ax, fig)

# displays combine figure ( heat map and image together)
plt.show()

def plot_heatmap_image(doc_id, vec1, vec2, url, text, model):
    # doc_id : index of the title1
    # vec1 : input apparels's vector, it is of a dict type {word:count}
    # vec2 : recommended apparels's vector, it is of a dict type {word:count}
    # url : apparels image url
    # text: title of recomonded apparel (used to keep title of image)

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
# 3. idf
```

```

# we find the common words in both titles, because these only words contribute to the dis
intersection = set(vec1.keys()) & set(vec2.keys())

# we set the values of non intersecting words to zero, this is just to show the differenc
for i in vec2:
    if i not in intersection:
        vec2[i]=0

# for labeling heatmap, keys contains list of all words in title2
keys = list(vec2.keys())
# if ith word in intersection(lis of words of title1 and list of words of title2): value
values = [vec2[x] for x in vec2.keys()]

# labels: len(labels) == len(keys), the values of labels depends on the model we are usin
# if model == 'bag of words': labels(i) = values(i)
# if model == 'tfidf weighted bag of words':labels(i) = tfidf(keys(i))
# if model == 'idf weighted bag of words':labels(i) = idf(keys(i))

```

```

if model == 'bag_of_words':
    labels = values
elif model == 'tfidf':
    labels = []
    for x in vec2.keys():
        # tfidf_title_vectorizer.vocabulary_ it contains all the words in the corpus
        # tfidf_title_features[doc_id, index_of_word_in_corpus] will give the tfidf value
        if x in tfidf_title_vectorizer.vocabulary_:
            labels.append(tfidf_title_features[doc_id, tfidf_title_vectorizer.vocabulary_[x]])
        else:
            labels.append(0)
elif model == 'idf':
    labels = []
    for x in vec2.keys():
        # idf_title_vectorizer.vocabulary_ it contains all the words in the corpus
        # idf_title_features[doc_id, index_of_word_in_corpus] will give the idf value of
        if x in idf_title_vectorizer.vocabulary_:
            labels.append(idf_title_features[doc_id, idf_title_vectorizer.vocabulary_[x]])
        else:
            labels.append(0)

plot_heatmap(keys, values, labels, url, text)

```

```

# this function gets a list of words along with the frequency of each
# word given "text"
def text_to_vector(text):
    word = re.compile(r'\w+')
    words = word.findall(text)

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

```

def get_result(doc_id, content_a, content_b, url, model):
    text1 = content_a
    text2 = content_b

    # vector1 = dict{word11:#count, word12:#count, etc.}
    vector1 = text_to_vector(text1)

    # vector2 = dict{word21:#count, word22:#count, etc.}
    vector2 = text_to_vector(text2)

    plot_heatmap_image(doc_id, vector1, vector2, url, text2, model)

```

▼ [8.2] Bag of Words (BoW) on product titles.

```

from sklearn.feature_extraction.text import CountVectorizer

```

```
title_vectorizer = CountVectorizer()
title_features = title_vectorizer.fit_transform(data['title'])
title_features.get_shape() # get number of rows and columns in feature matrix.
```

 (16042, 12609)

```
def bag_of_words_model(doc_id, num_results):
    # doc_id: apparel's id in given corpus

    # pairwise_dist will store the distance from given input apparel to all remaining apparel
    # the metric we used here is cosine, the cosine distance is measured as  $K(X, Y) = \langle X, Y \rangle / \sqrt{\langle X, X \rangle \langle Y, Y \rangle}$ 
    # http://scikit-learn.org/stable/modules/generated/sklearn.metrics.pairwise.cosine\_similarity.html
    pairwise_dist = pairwise_distances(title_features, title_features[doc_id])

    # np.argsort will return indices of the smallest distances
    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    # pdists will store the smallest distances
    pdists = np.sort(pairwise_dist.flatten())[0:num_results]

    # data frame indices of the 9 smallest distance's
    df_indices = list(data.index[indices])

    for i in range(0, len(indices)):
        # we will pass 1. doc_id, 2. title1, 3. title2, url, model
        get_result(indices[i], data['title'].loc[df_indices[0]], data['title'].loc[df_indices[i]])
        print('ASIN :', data['asin'].loc[df_indices[i]])
        print('Brand:', data['brand'].loc[df_indices[i]])
        print('Title:', data['title'].loc[df_indices[i]])
        print('Euclidean similarity with the query image :', pdists[i])
```

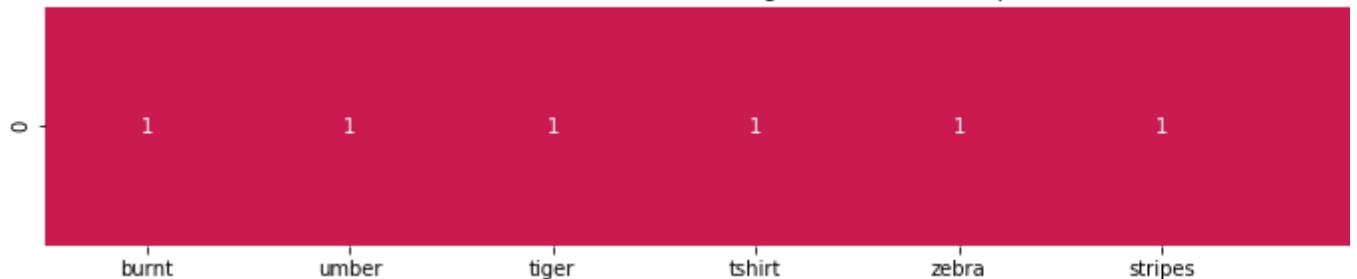
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) 

```
# call the bag-of-words model for a product to get similar products.
bag_of_words_model(12566, 20) # change the index if you want to.
# In the output heat map each value represents the count value
# of the label word, the color represents the intersection
# with inputs title.
```

```
#try 12566
#try 931
```



burnt umber tiger tshirt zebra stripes xl xxl



ASIN : B00JXQB5FQ

Brand: Si Row

Title: burnt umber tiger tshirt zebra stripes xl xxl

Euclidean similarity with the query image : 0.0

=====

pink tiger tshirt zebra stripes xl xxl



ASIN : B00JXQASS6

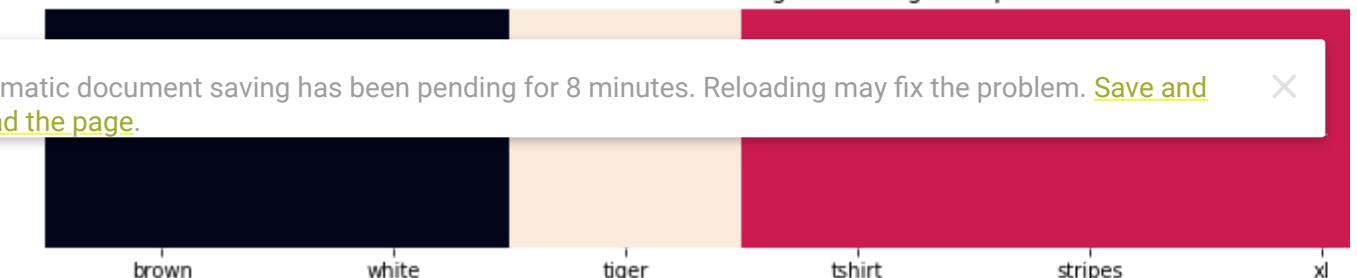
Brand: Si Row

Title: pink tiger tshirt zebra stripes xl xxl

Euclidean similarity with the query image : 1.7320508075688772

=====

brown white tiger tshirt tiger stripes xl xxl



ASIN : B00JXQCWT0

Brand: Si Row

Title: brown white tiger tshirt tiger stripes xl xxl

Euclidean similarity with the query image : 2.449489742783178

=====

yellow tiger tshirt tiger stripes l



ASIN : B00JXQCUIC

Brand: Si Row

Title: yellow tiger tshirt tiger stripes l

Euclidean similarity with the query image : 2.6457513110645907

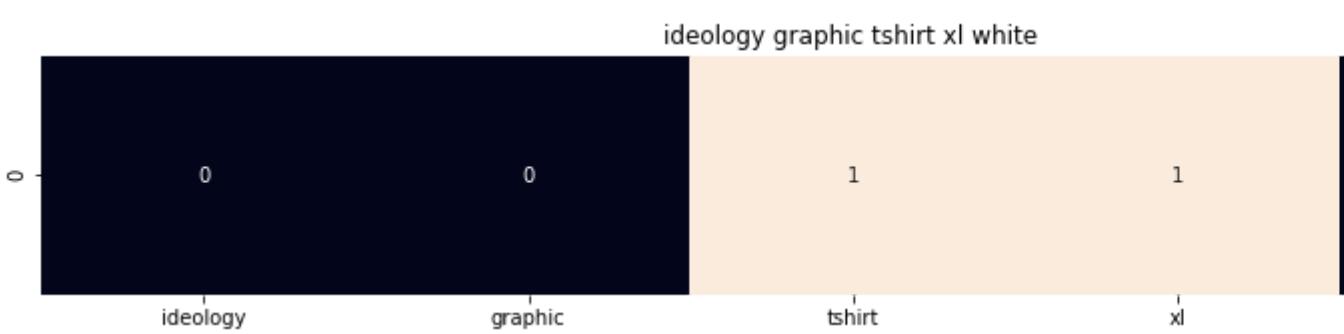


ASIN : B07568NZX4

Brand: Rustic Grace

Title: believed could tshirt

Euclidean similarity with the query image : 3.0

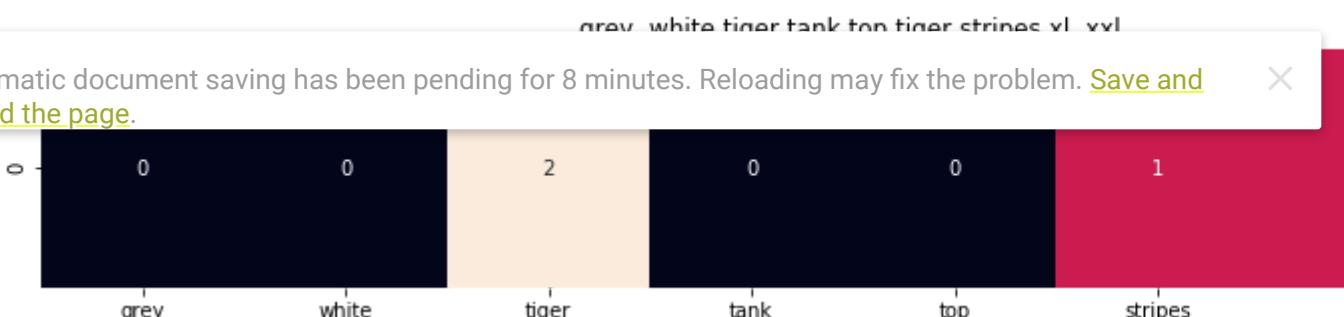


ASIN : B01NB0NKRO

Brand: Ideology

Title: ideology graphic tshirt xl white

Euclidean similarity with the query image : 3.0



ASIN : B00JXQAFZ2

Brand: Si Row

Title: grey white tiger tank top tiger stripes xl xxl

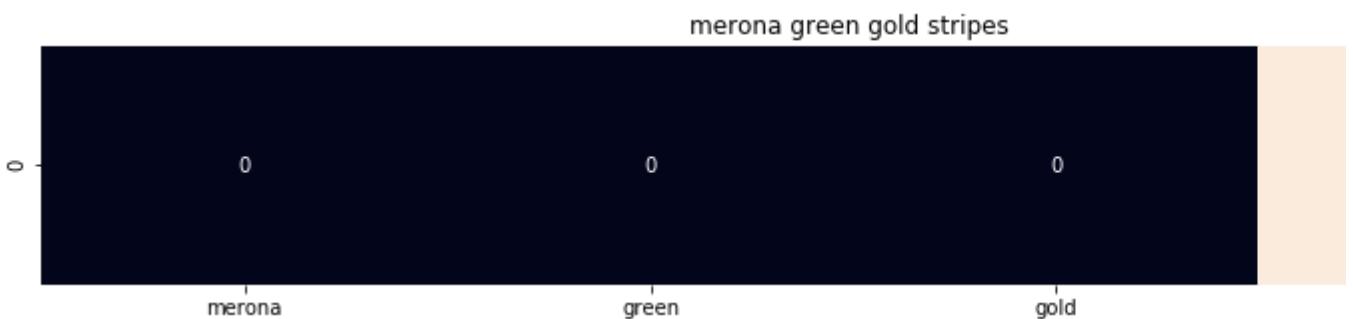
Euclidean similarity with the query image : 3.0



ASIN : B01CLS8LMW

Brand: Awake

Title: morning person tshirt troll picture xl
 Euclidean similarity with the query image : 3.1622776601683795

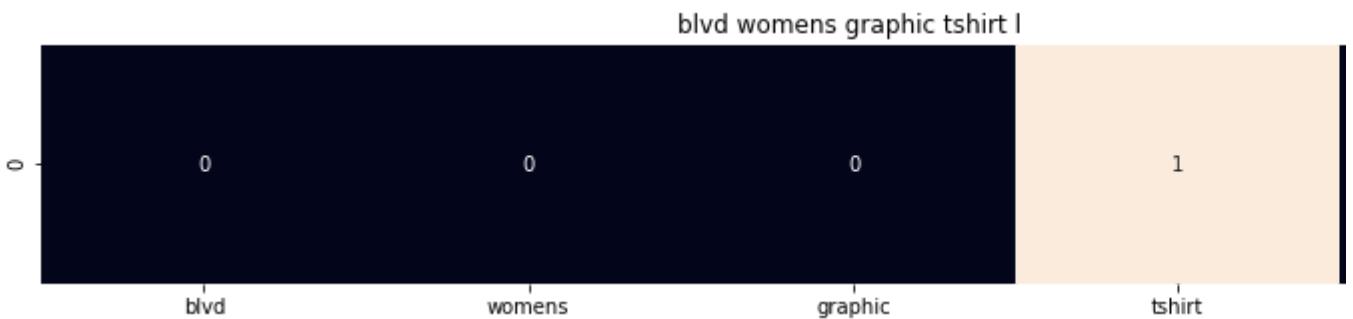


ASIN : B01KVZUB6G

Brand: Merona

Title: merona green gold stripes

Euclidean similarity with the query image : 3.1622776601683795



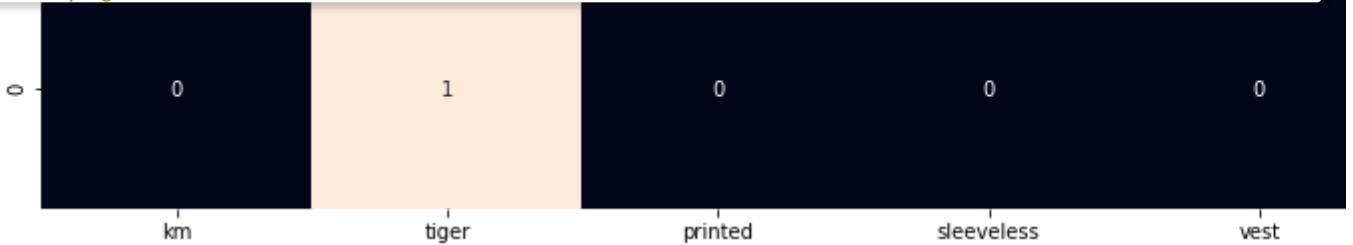
ASIN : B0733R2CJK

Brand: BLVD

Title: blvd womens graphic tshirt l

Euclidean similarity with the query image : 3.1622776601683795

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

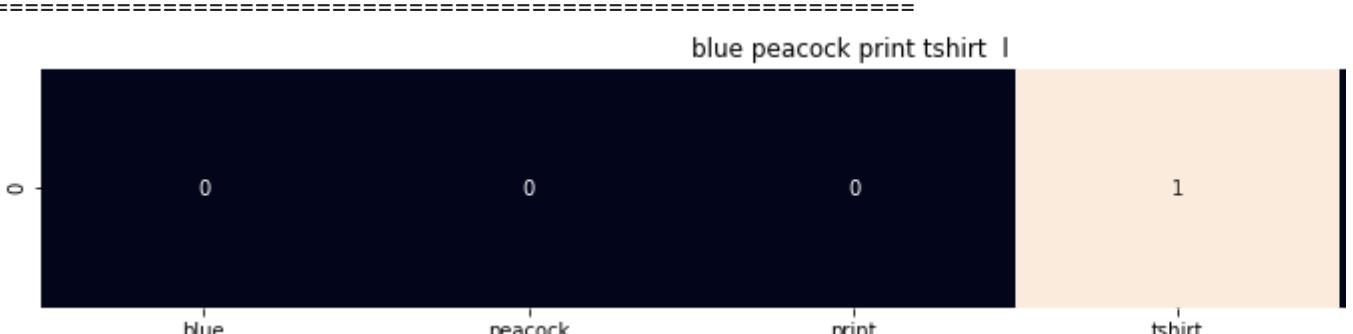


ASIN : B012VQLT6Y

Brand: KM T-shirt

Title: km tiger printed sleeveless vest tshirt

Euclidean similarity with the query image : 3.1622776601683795



ASIN : B000JXQCL8L9

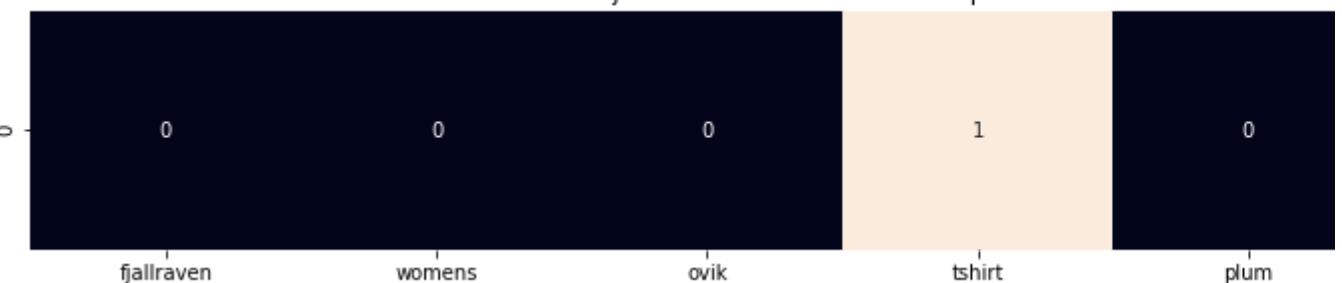
Brand: Si Row

Title: blue peacock print tshirt 1

Euclidean similarity with the query image : 3.1622776601683795

=====

fjallraven womens ovik tshirt plum xxl



ASIN : B06XC3CZF6

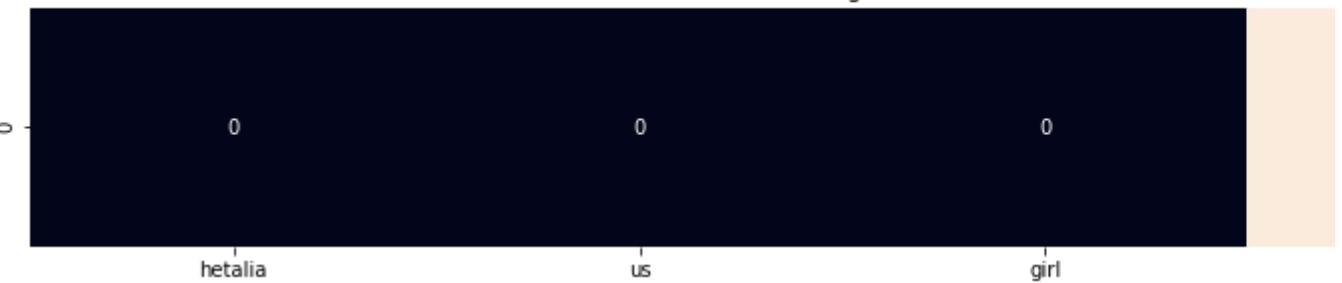
Brand: Fjallraven

Title: fjallraven womens ovik tshirt plum xxl

Euclidean similarity with the query image : 3.1622776601683795

=====

hetalia us girl tshirt

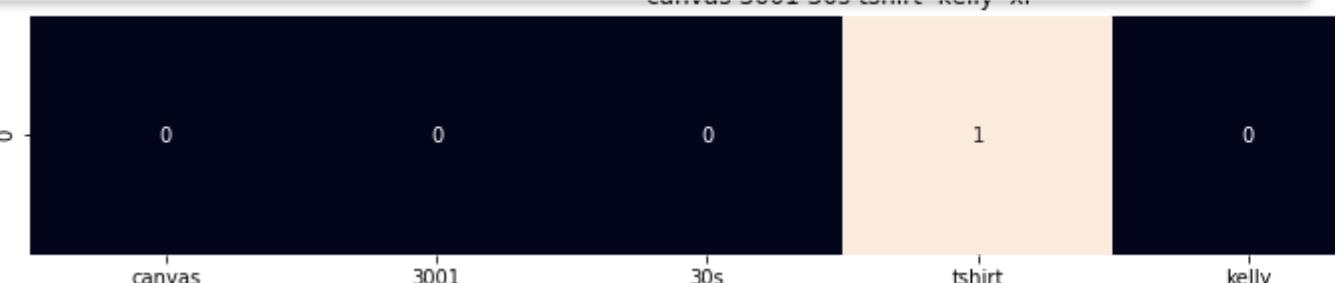


ASIN : B005IT80BA

Brand: Hetalia

Title: hetalia us girl tshirt

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).



ASIN : B0088PN0LA

Brand: Red House

Title: canvas 3001 30s tshirt kelly xl

Euclidean similarity with the query image : 3.1622776601683795

=====

brunello cucinelli tshirt women white xl



brunello

cucinelli

tshirt

women

white

ASIN : B06X99V6WC

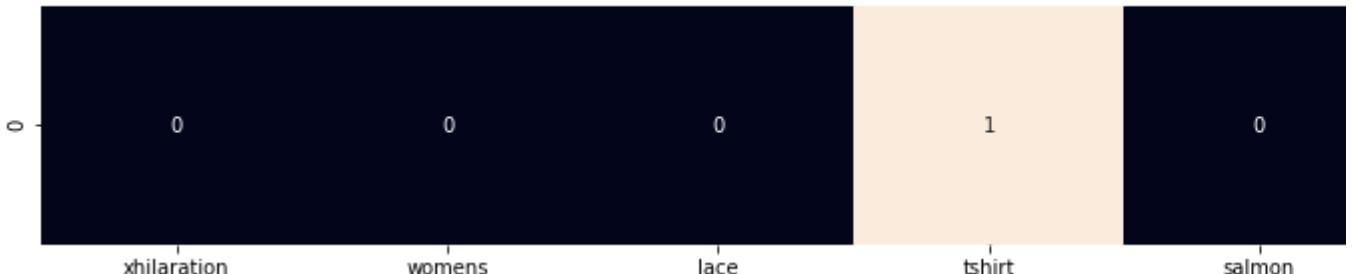
Brand: Brunello Cucinelli

Title: brunello cucinelli tshirt women white xl

Euclidean similarity with the query image : 3.1622776601683795

=====

xhilaration womens lace tshirt salmon xxl



ASIN : B06Y1JPW1Q

Brand: Xhilaration

Title: xhilaration womens lace tshirt salmon xxl

Euclidean similarity with the query image : 3.1622776601683795

=====

animal oceania tshirt yellow



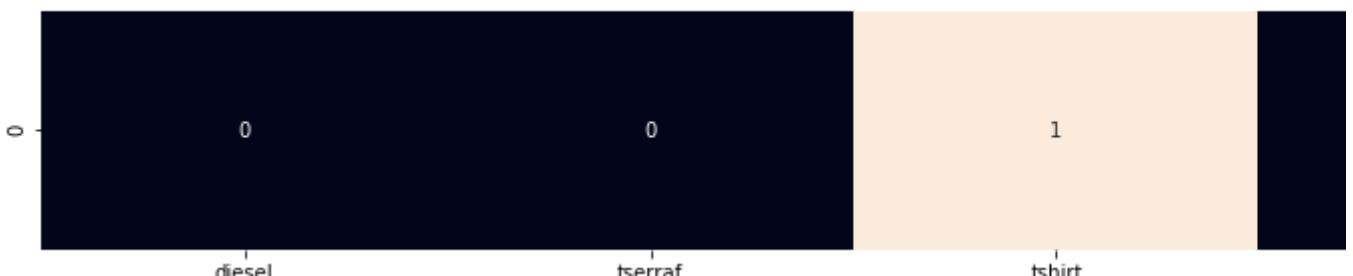
ASIN : B06X6GX6WG

Brand: Animal

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#). X

=====

diesel tserraf tshirt black



ASIN : B017X8PW9U

Brand: Diesel

Title: diesel tserraf tshirt black

Euclidean similarity with the query image : 3.1622776601683795

=====

juniors love lucyaaaaahhhh tshirt size xl



juniors	love	lucywaaaaahhhh	tshirt	size
---------	------	----------------	--------	------

ASIN : B00IAAA4JIQ
 Brand: I Love Lucy
 Title: juniors love lucywaaaaahhhh tshirt size xl
 Euclidean similarity with the query image : 3.1622776601683795
 =====

▼ [8.5] TF-IDF based product similarity

```
tfidf_title_vectorizer = TfidfVectorizer(min_df = 0)
tfidf_title_features = tfidf_title_vectorizer.fit_transform(data['title'])

def tfidf_model(doc_id, num_results):
    # doc_id: apparel's id in given corpus

    # pairwise_dist will store the distance from given input apparel to all remaining apparel
    # the metric we used here is cosine, the cosine distance is measured as K(X, Y) = <X, Y> /
    # http://scikit-learn.org/stable/modules/metrics.html#cosine-similarity
    pairwise_dist = pairwise_distances(tfidf_title_features,tfidf_title_features[doc_id])

    # np.argsort will return indices of 9 smallest distances
    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    #pdists will store the 9 smallest distances
    pdists = pairwise_dist.flatten()[0:num_results]

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. Save and reload the page X
```

```
df_indices = list(data.index[indices])

for i in range(0,len(indices)):
    # we will pass 1. doc_id, 2. title1, 3. title2, url, model
    get_result(indices[i], data['title'].loc[df_indices[0]], data['title'].loc[df_indices[1]])
    print('ASIN :',data['asin'].loc[df_indices[i]])
    print('BRAND :',data['brand'].loc[df_indices[i]])
    print ('Euclidean distance from the given image :', pdists[i])
    print('*'*125)
tfidf_model(12566, 20)
# in the output heat map each value represents the tfidf values of the label word, the color
```



burnt umber tiger tshirt zebra stripes xl xxl



ASIN : B00JXQB5FQ

BRAND : Si Row

Eucliden distance from the given image : 0.0

pink tiger tshirt zebra stripes xl xxl

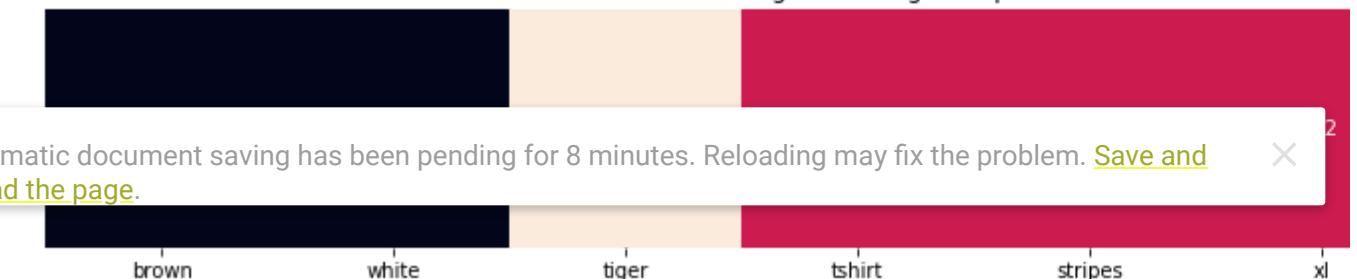


ASIN : B00JXQASS6

BRAND : Si Row

Eucliden distance from the given image : 0.7536331912451361

brown white tiger tshirt tiger stripes xl xxl



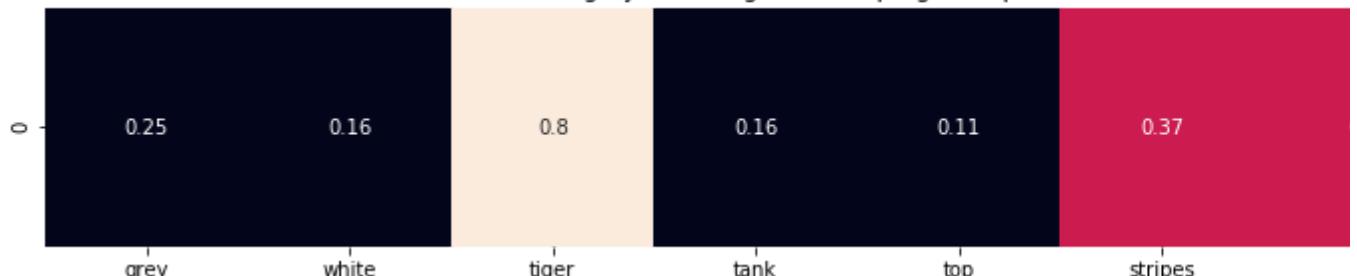
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQCWT0

BRAND : Si Row

Eucliden distance from the given image : 0.9357643943769645

grey white tiger tank top tiger stripes xl xxl



ASIN : B00JXQAFZ2

BRAND : Si Row

Eucliden distance from the given image : 0.9586153524200749

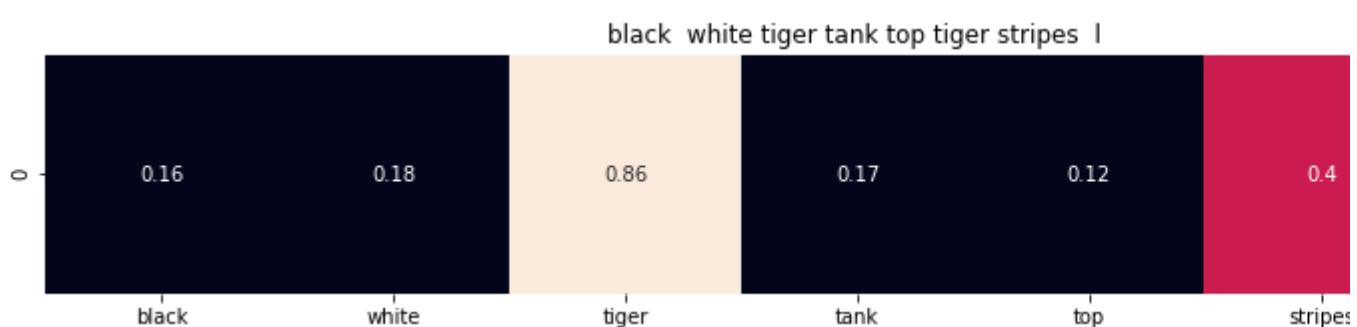
yellow tiger tshirt tiger stripes l



ASIN : B00JXQCUIC

BRAND : Si Row

Eucliden distance from the given image : 1.000074961446881



ASIN : B00JXQA094

BRAND : Si Row

Eucliden distance from the given image : 1.023215552457452

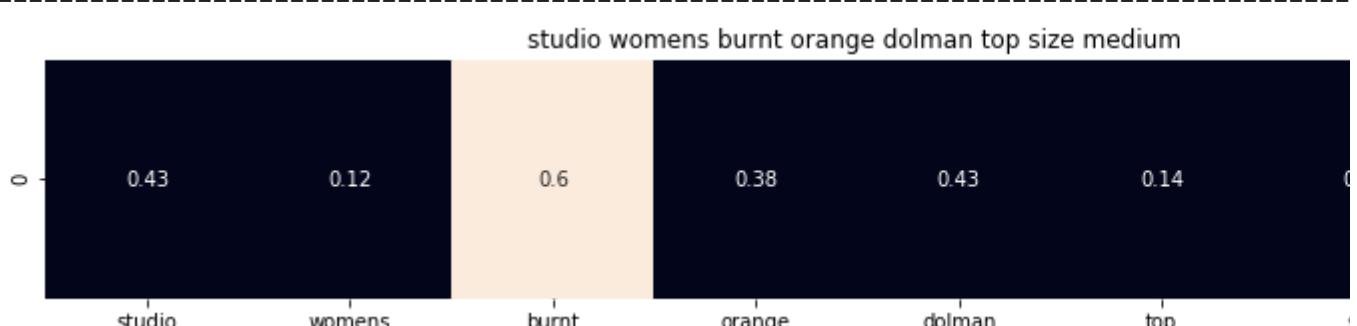


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

ASIN : B00JXQAUWA

BRAND : Si Row

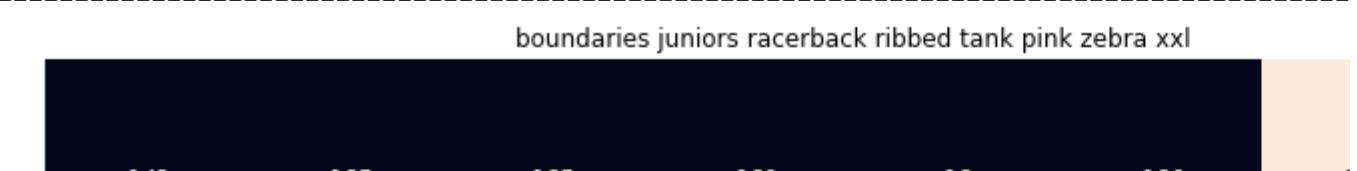
Eucliden distance from the given image : 1.031991846303421



ASIN : B06XSCVFT5

BRAND : Studio M

Eucliden distance from the given image : 1.2106843670424716

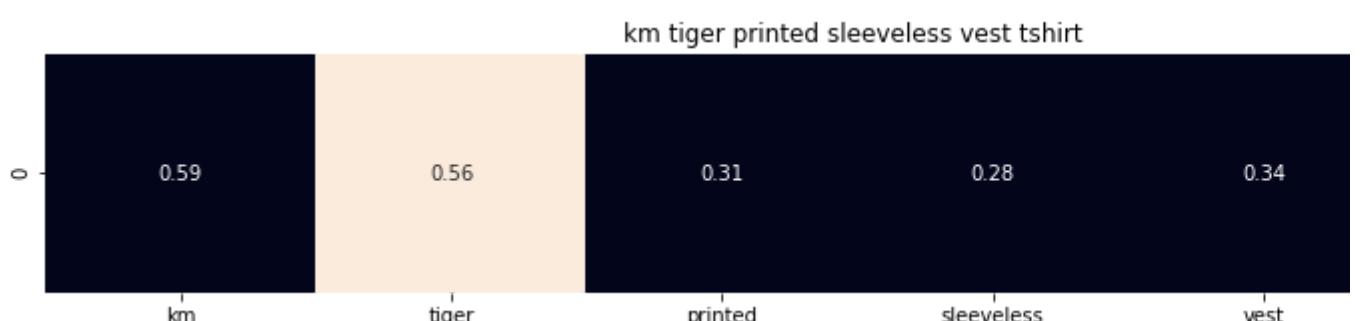




ASIN : B06Y2GTYPM

BRAND : No Boundaries

Euclidean distance from the given image : 1.212168381072083



ASIN : B012VQLT6Y

BRAND : KM T-shirt

Euclidean distance from the given image : 1.219790640280982



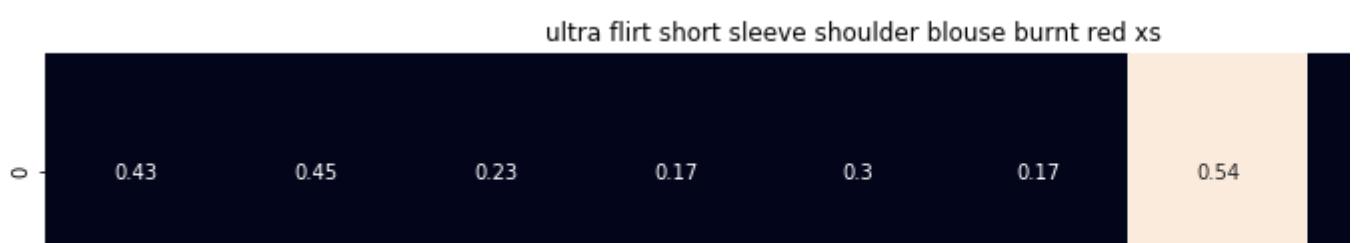
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).



ASIN : B00Z6HEXWI

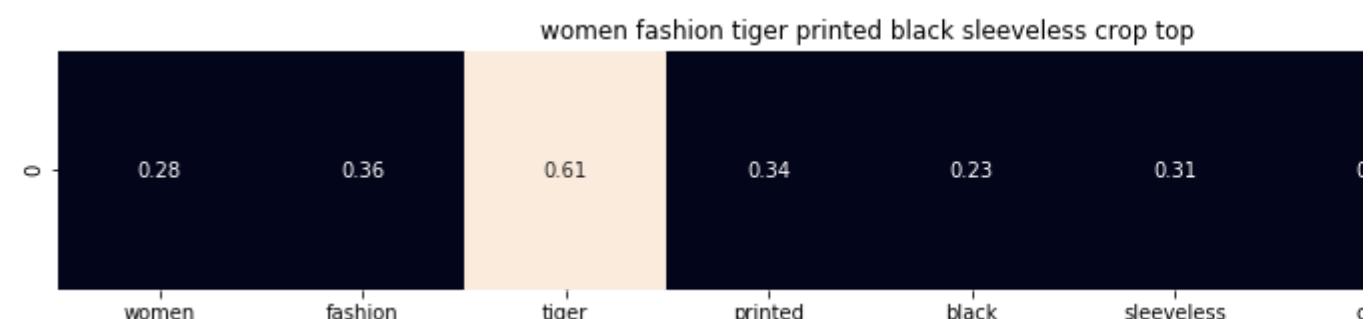
BRAND : Black Temptation

Euclidean distance from the given image : 1.221281392120943



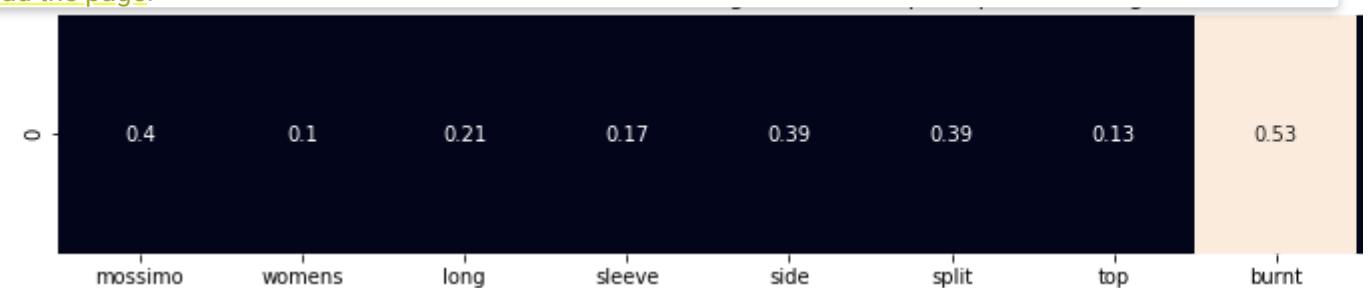


ASIN : B072R2JXKW
 BRAND : WHAT ON EARTH
 Euclidean distance from the given image : 1.2318451972624518



ASIN : B074T8ZYGX
 BRAND : MKP Crop Top

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).



ASIN : B071ZDF6T2
 BRAND : Mossimo
 Euclidean distance from the given image : 1.2352785577664824

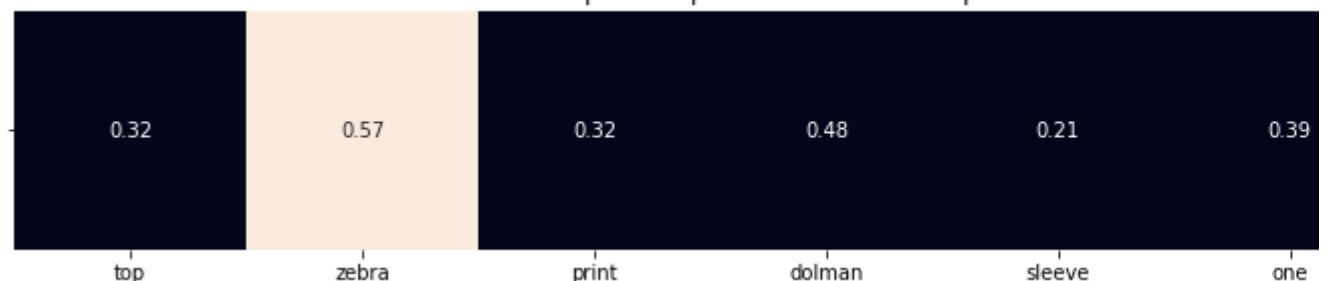


ASIN : B01K0H020G

BRAND : Tultex

Eucliden distance from the given image : 1.236457298812782

===== top zebra print dolman sleeve top one size

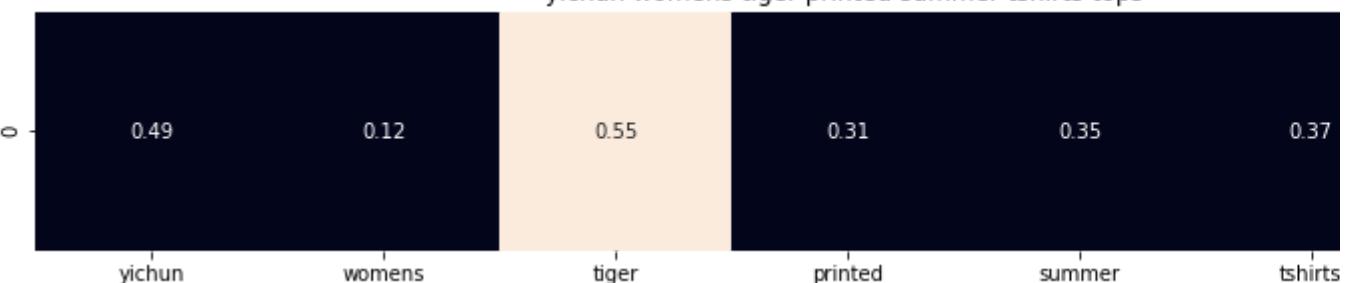


ASIN : B00H8A6ZLI

BRAND : Vivian's Fashions

Eucliden distance from the given image : 1.24996155052848

===== yichun womens tiger printed summer tshirts tops



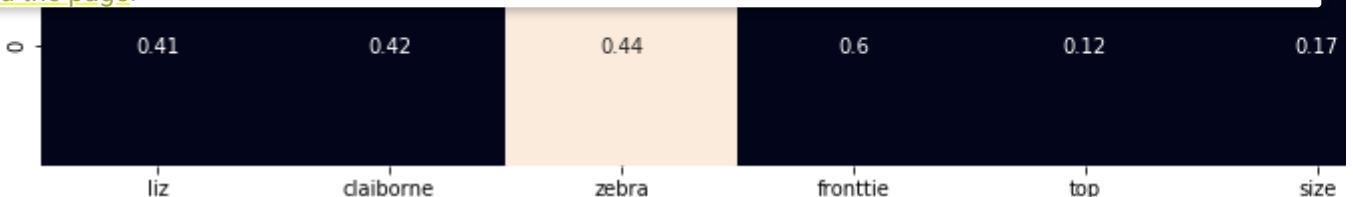
ASIN : B010NN9RX0

BRAND : YICHUN

Eucliden distance from the given image : 1.25354614208561

===== liz claiborne zebra fronttie top size xl

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B06XBY5QXL

BRAND : Liz Claiborne

Eucliden distance from the given image : 1.2538832938357722

▼ [8.5] IDF based product similarity

```
idf_title_vectorizer = CountVectorizer()
idf_title_features = idf_title_vectorizer.fit_transform(data['title'])
```

```
def n_containing(word):
```

https://colab.research.google.com/drive/1MAuRSmm3p154U3ZYXZpF_E0EJeAE6nS1#scrollTo=u5CQeoZN5Y-&printMode=true

```

nContaining(word):
    # return the number of documents which had the given word
    return sum(1 for blob in data['title'] if word in blob.split())

def idf(word):
    # idf = log(#number of docs / #number of docs which had the given word)
    return math.log(data.shape[0] / (nContaining(word)))

# we need to convert the values into float
idf_title_features = idf_title_features.astype(np.float)

for i in idf_title_vectorizer.vocabulary_.keys():
    # for every word in whole corpus we will find its idf value
    idf_val = idf(i)

    # to calculate idf_title_features we need to replace the count values with the idf values
    # idf_title_features[:, idf_title_vectorizer.vocabulary_[i]].nonzero()[0] will return all
    for j in idf_title_features[:, idf_title_vectorizer.vocabulary_[i]].nonzero()[0]:
        # we replace the count values of word i in document j with idf_value of word i
        # idf_title_features[doc_id, index_of_word_in_corpus] = idf value of word
        idf_title_features[j,idf_title_vectorizer.vocabulary_[i]] = idf_val

```

```

def idf_model(doc_id, num_results):
    # doc_id: apparel's id in given corpus

    # pairwise_dist will store the distance from given input apparel to all remaining apparel

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```

# np.argsort will return indices of 9 smallest distances
indices = np.argsort(pairwise_dist.flatten())[0:num_results]
#pdists will store the 9 smallest distances
pdists = np.sort(pairwise_dist.flatten())[0:num_results]

#data frame indices of the 9 smallest distace's
df_indices = list(data.index[indices])

for i in range(0,len(indices)):
    get_result(indices[i],data['title'].loc[df_indices[0]], data['title'].loc[df_indices[1]])
    print('ASIN :',data['asin'].loc[df_indices[i]])
    print('Brand :',data['brand'].loc[df_indices[i]])
    print ('euclidean distance from the given image :', pdists[i])
    print('*125)

```

```
idf_model(12566,20)
```

```
# in the output heat map each value represents the idf values of the label word, the color re
```



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

burnt umber tiger tshirt zebra stripes xl xxl

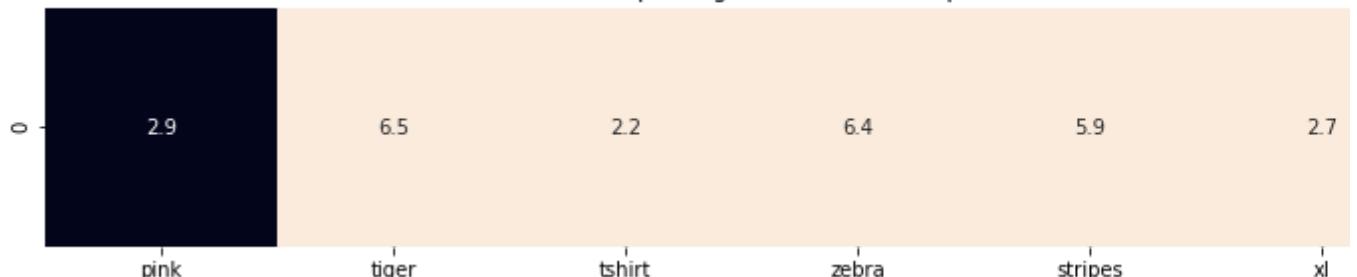


ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from the given image : 0.0

pink tiger tshirt zebra stripes xl xxl

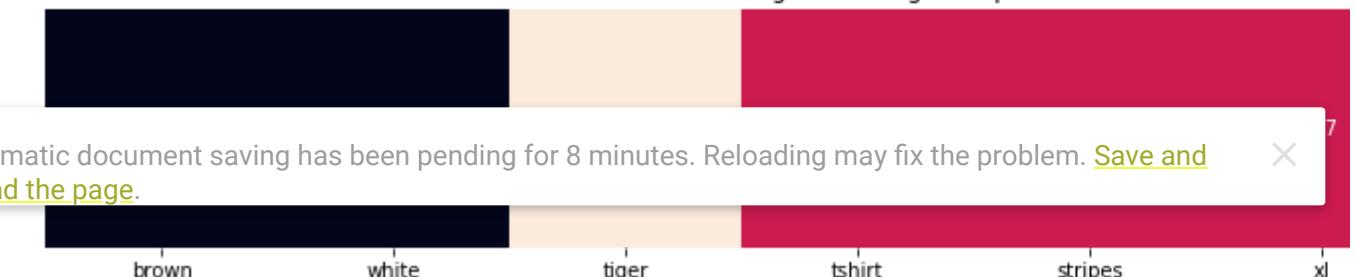


ASIN : B00JXQASS6

Brand : Si Row

euclidean distance from the given image : 12.20507131122177

brown white tiger tshirt tiger stripes xl xxl

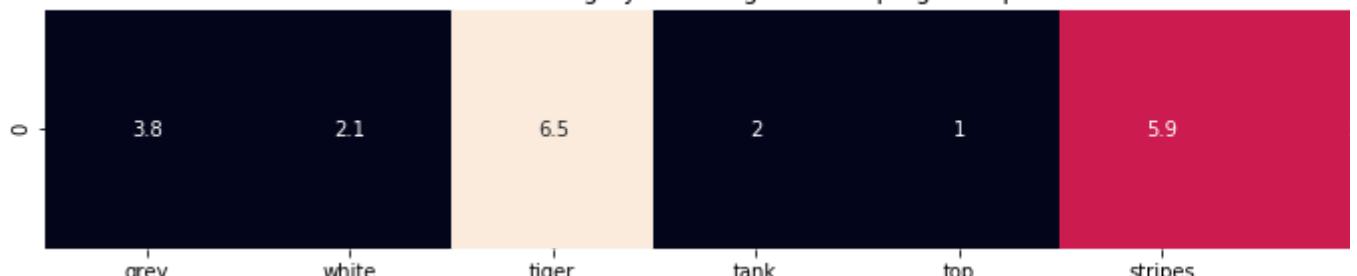
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQCWT0

Brand : Si Row

euclidean distance from the given image : 14.468362685603465

grey white tiger tank top tiger stripes xl xxl

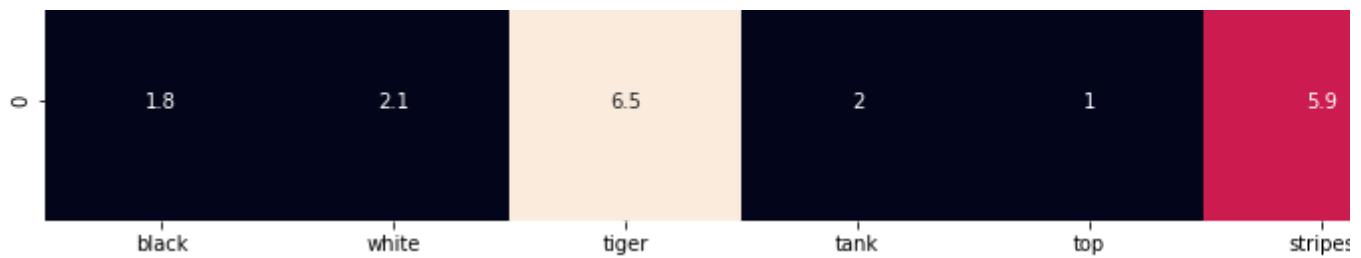


ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from the given image : 14.486832924778964

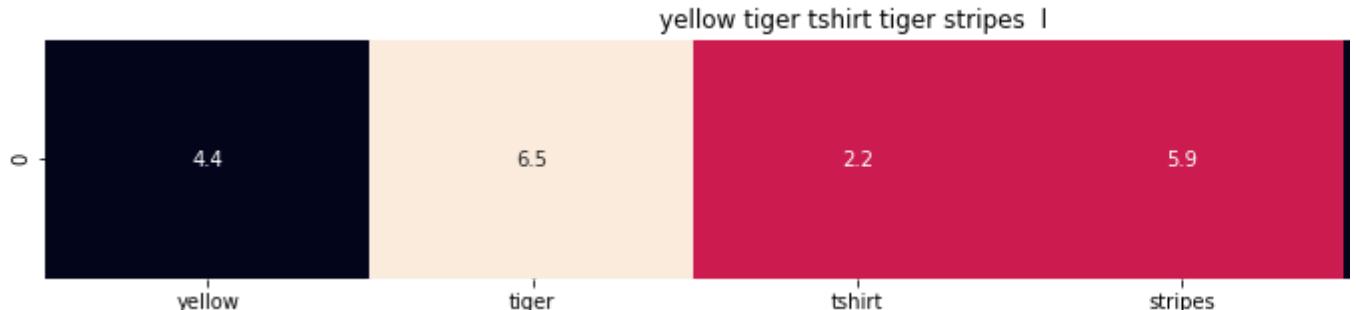
black white tiger tank top tiger stripes l



ASIN : B00JXQA094

Brand : Si Row

euclidean distance from the given image : 14.833392966672909



ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from the given image : 14.898744516719225

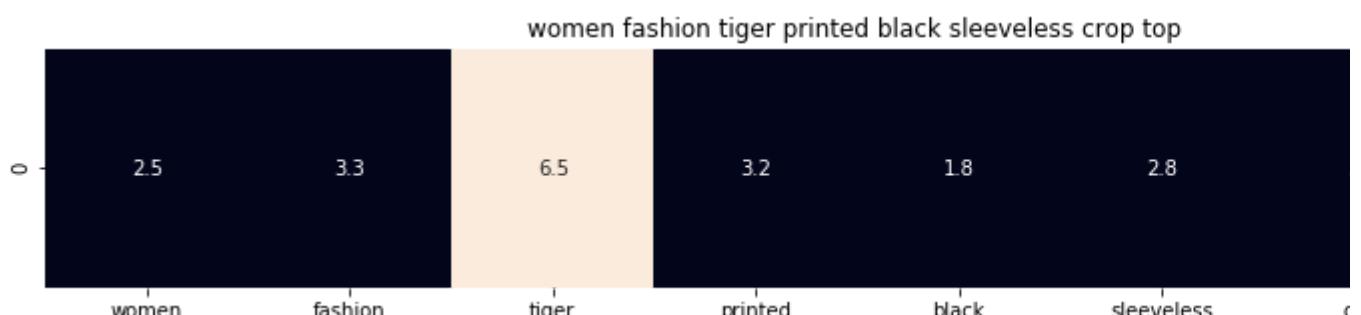


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

ASIN : B00JXQUWA

Brand : Si Row

euclidean distance from the given image : 15.224458287343769



ASIN : B074T8ZYGX

Brand : MKP Crop Top

euclidean distance from the given image : 17.080812955631995

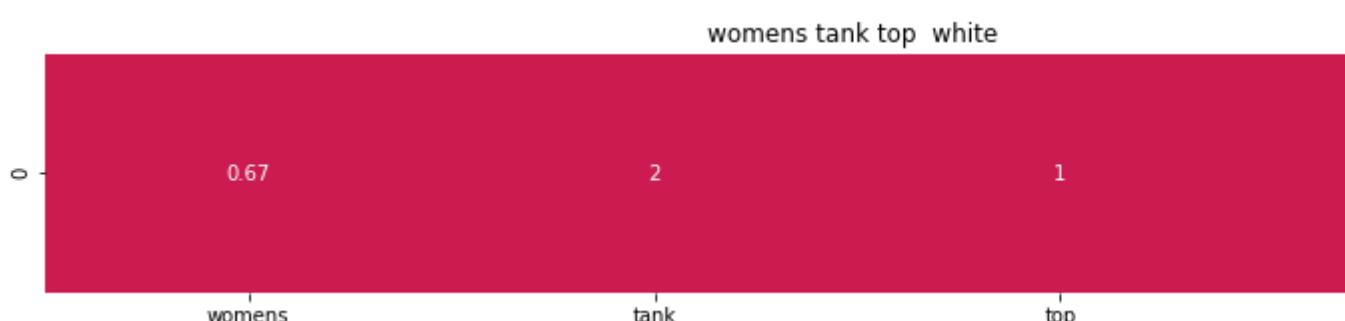




ASIN : B00KF2N5PU

Brand : Vietsbay

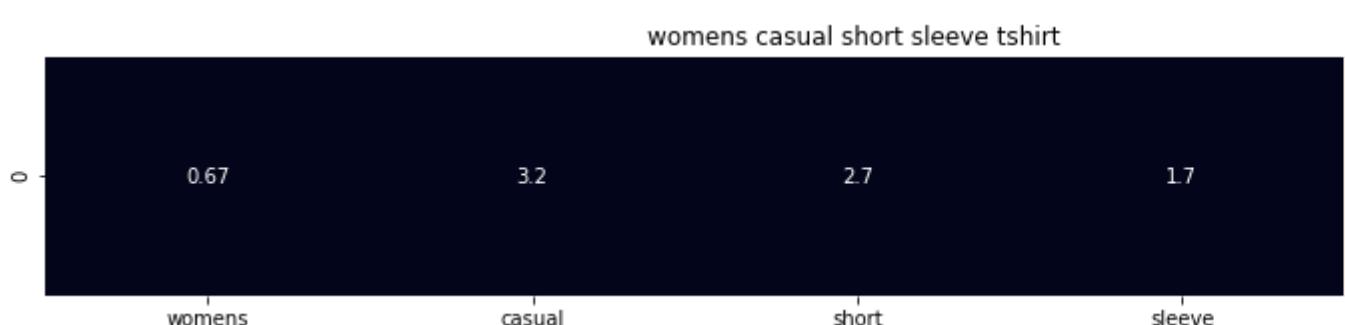
euclidean distance from the given image : 17.090168125645416



ASIN : B00JP0Z9GM

Brand : Sofra

euclidean distance from the given image : 17.153215337562703



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and](#)

[reload the page](#).

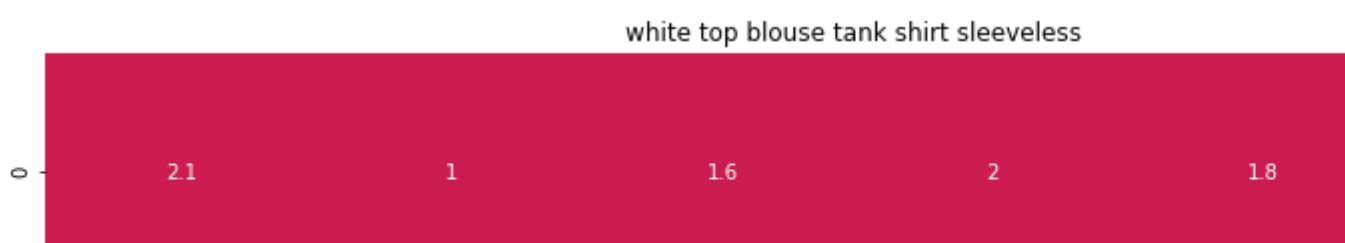
euclidean distance from the given image : 17.35071925074909



ASIN : B00H8A6ZLI

Brand : Vivian's Fashions

euclidean distance from the given image : 17.410075941001253



white

top

blouse

tank

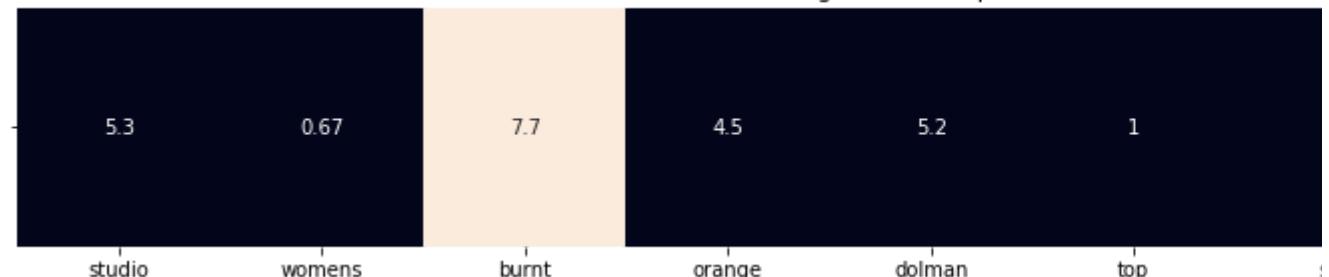
shirt

ASIN : B074G5G5RK

Brand : ERMANNO SCERVINO

euclidean distance from the given image : 17.539921335459557

===== studio womens burnt orange dolman top size medium =====



ASIN : B06XSCVFT5

Brand : Studio M

euclidean distance from the given image : 17.61275854366134

===== wear womens vneck blouse black xxl =====



ASIN : B06Y6FH453

Brand : Who What Wear

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X ==



ASIN : B074V45DCX

Brand : Rain

euclidean distance from the given image : 17.634342496835046

===== i women shoulder blue small =====

▼ [9] Text Semantics based product similarity

=====

```
# credits: https://www.kaggle.com/c/word2vec-nlp-tutorial#part-2-word-vectors
# Custom Word2Vec using your own text data.
# Do NOT RUN this code
```

```
# DO NOT RUN THIS CODE.  
# It is meant as a reference to build your own Word2Vec when you have  
# lots of data.
```

```
...  
  
# Set values for various parameters  
num_features = 300      # Word vector dimensionality  
min_word_count = 1       # Minimum word count  
num_workers = 4           # Number of threads to run in parallel  
context = 10              # Context window size  
downsampling = 1e-3       # Downsample setting for frequent words
```

```
# Initialize and train the model (this will take some time)  
from gensim.models import word2vec  
print ("Training model...")  
model = word2vec.Word2Vec(sen_corpus, workers=num_workers, \  
                         size=num_features, min_count = min_word_count, \  
                         window = context)
```

```
...
```

```
👤 '\n# Set values for various parameters\nnum_features = 300      # Word vector dimensionali
```

```
from gensim.models import Word2Vec  
from gensim.models import KeyedVectors  
import pickle
```

```
# in this project we are using a pretrained model by google
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
...  
# we will provide a pickle file with contains a dict ,  
# and it contains all our corpus words as keys and model[word] as values  
# To use this code-snippet, download "GoogleNews-vectors-negative300.bin"  
# from https://drive.google.com/file/d/0B7XkCwpI5KDYNlNUTT1SS21pQmM/edit  
# it's 1.9GB in size.
```

```
...
```

```
model = KeyedVectors.load_word2vec_format('GoogleNews-vectors-negative300.bin', binary=True)  
...
```

```
#if you do NOT have RAM >= 12GB, use the code below.  
with open('word2vec_model', 'rb') as handle:  
    model = pickle.load(handle)
```

```
# Utility functions
```

```
def get_word_vec(sentence, doc_id, m_name):  
    # sentence : title of the apparel  
    # doc_id: document id in our corpus  
    ...
```

```

# m_name: model information it will take two values
# if m_name == 'avg', we will append the model[i], w2v representation of word i
# if m_name == 'weighted', we will multiply each w2v[word] with the idf(word)
vec = []
for i in sentence.split():
    if i in vocab:
        if m_name == 'weighted' and i in idf_title_vectorizer.vocabulary_:
            vec.append(idf_title_features[doc_id, idf_title_vectorizer.vocabulary_[i]] *
        elif m_name == 'avg':
            vec.append(model[i])
    else:
        # if the word in our courpus is not there in the google word2vec corpus, we are j
        vec.append(np.zeros(shape=(300,)))
# we will return a numpy array of shape (#number of words in title * 300 ) 300 = len(w2v_
# each row represents the word2vec representation of each word (weighted/avg) in given se
return np.array(vec)

def get_distance(vec1, vec2):
    # vec1 = np.array(#number_of_words_title1 * 300), each row is a vector of length 300 corr
    # vec2 = np.array(#number_of_words_title2 * 300), each row is a vector of length 300 corr

    final_dist = []
    # for each vector in vec1 we caluclate the distance(euclidean) to all vectors in vec2
    for i in vec1:
        dist = []
        for j in vec2:
            # np.linalg.norm(i-j) will result the euclidean distance between vectors i, j
            dist.append(np.linalg.norm(i-j))
        final_dist.append(np.array(dist))

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```

def heat_map_w2v(sentence1, sentence2, url, doc_id1, doc_id2, model):
    # sentance1 : title1, input apparel
    # sentance2 : title2, recommended apparel
    # url: apparel image url
    # doc_id1: document id of input apparel
    # doc_id2: document id of recommended apparel
    # model: it can have two values, 1. avg 2. weighted

    #s1_vec = np.array(#number_of_words_title1 * 300), each row is a vector(weighted/avg) of
    s1_vec = get_word_vec(sentence1, doc_id1, model)
    #s2_vec = np.array(#number_of_words_title1 * 300), each row is a vector(weighted/avg) of
    s2_vec = get_word_vec(sentence2, doc_id2, model)

    # s1_s2_dist = np.array(#number of words in title1 * #number of words in title2)
    # s1_s2_dist[i,j] = euclidean distance between words i, j
    s1_s2_dist = get_distance(s1_vec, s2_vec)

```

```

# devide whole figure into 2 parts 1st part displays heatmap 2nd part displays image of a
gs = gridspec.GridSpec(2, 2, width_ratios=[4,1],height_ratios=[2,1])
fig = plt.figure(figsize=(15,15))

ax = plt.subplot(gs[0])
# ploting the heap map based on the pairwise distances
ax = sns.heatmap(np.round(s1_s2_dist,4), annot=True)
# set the x axis labels as recommended apparels title
ax.set_xticklabels(sentence2.split())
# set the y axis labels as input apparels title
ax.set_yticklabels(sentence1.split())
# set title as recommended apparels title
ax.set_title(sentence2)

ax = plt.subplot(gs[1])
# we remove all grids and axis labels for image
ax.grid(False)
ax.set_xticks([])
ax.set_yticks([])
display_img(url, ax, fig)

plt.show()

```

```

# vocab = stores all the words that are there in google w2v model
# vocab = model.wv.vocab.keys() # if you are using Google word2Vec

vocab = model.keys()

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```

# num_features: the lenght of word2vec vector, its values = 300
# m_name: model information it will take two values
    # if m_name == 'avg', we will append the model[i], w2v representation of word i
    # if m_name == 'weighted', we will multiply each w2v[word] with the idf(word)

featureVec = np.zeros((num_features,), dtype="float32")
# we will intialize a vector of size 300 with all zeros
# we add each word2vec(wordi) to this festureVec
nwords = 0

for word in sentence.split():
    nwords += 1
    if word in vocab:
        if m_name == 'weighted' and word in idf_title_vectorizer.vocabulary_:
            featureVec = np.add(featureVec, idf_title_features[doc_id, idf_title_vectorizer.vocabulary_[word]])
        elif m_name == 'avg':
            featureVec = np.add(featureVec, model[word])
    if(nwords>0):
        featureVec = np.divide(featureVec, nwords)
# returns the avg vector of given sentance, its of shape (1, 300)

```

```
return featureVec
```

▼ [9.2] Average Word2Vec product similarity.

```
doc_id = 0
w2v_title = []
# for every title we build a avg vector representation
for i in data['title']:
    w2v_title.append(build_avg_vec(i, 300, doc_id, 'avg'))
    doc_id += 1

# w2v_title = np.array(# number of doc in courpus * 300), each row corresponds to a doc
w2v_title = np.array(w2v_title)
```

```
def avg_w2v_model(doc_id, num_results):
    # doc_id: apparel's id in given corpus

    # dist(x, y) = sqrt(dot(x, x) - 2 * dot(x, y) + dot(y, y))
    pairwise_dist = pairwise_distances(w2v_title, w2v_title[doc_id].reshape(1,-1))

    # np.argsort will return indices of 9 smallest distances
    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    #pdists will store the 9 smallest distances
    pdists = np.sort(pairwise_dist.flatten())[0:num_results]
```

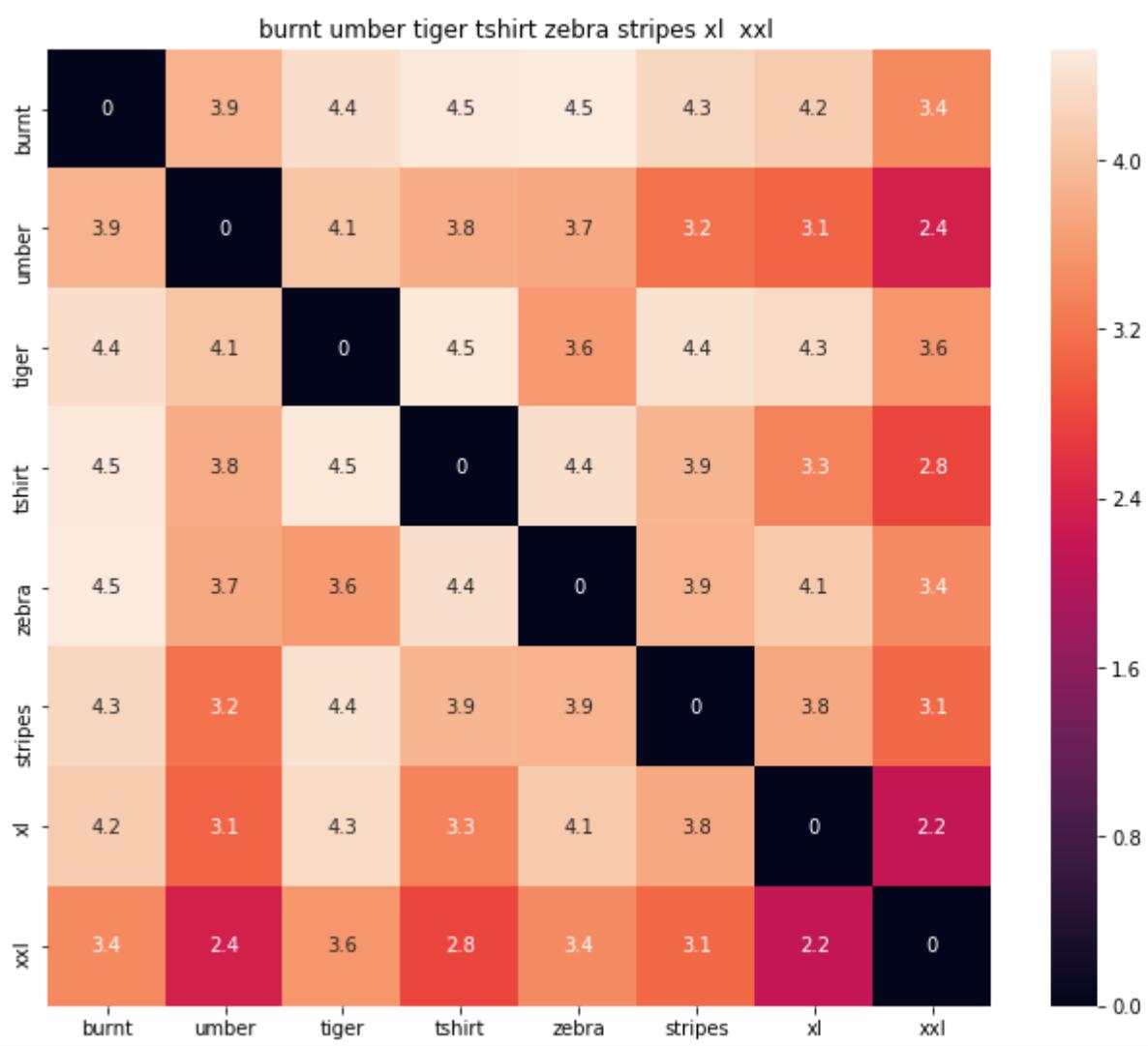
~~#data frame indices of the 9 smallest distances~~

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

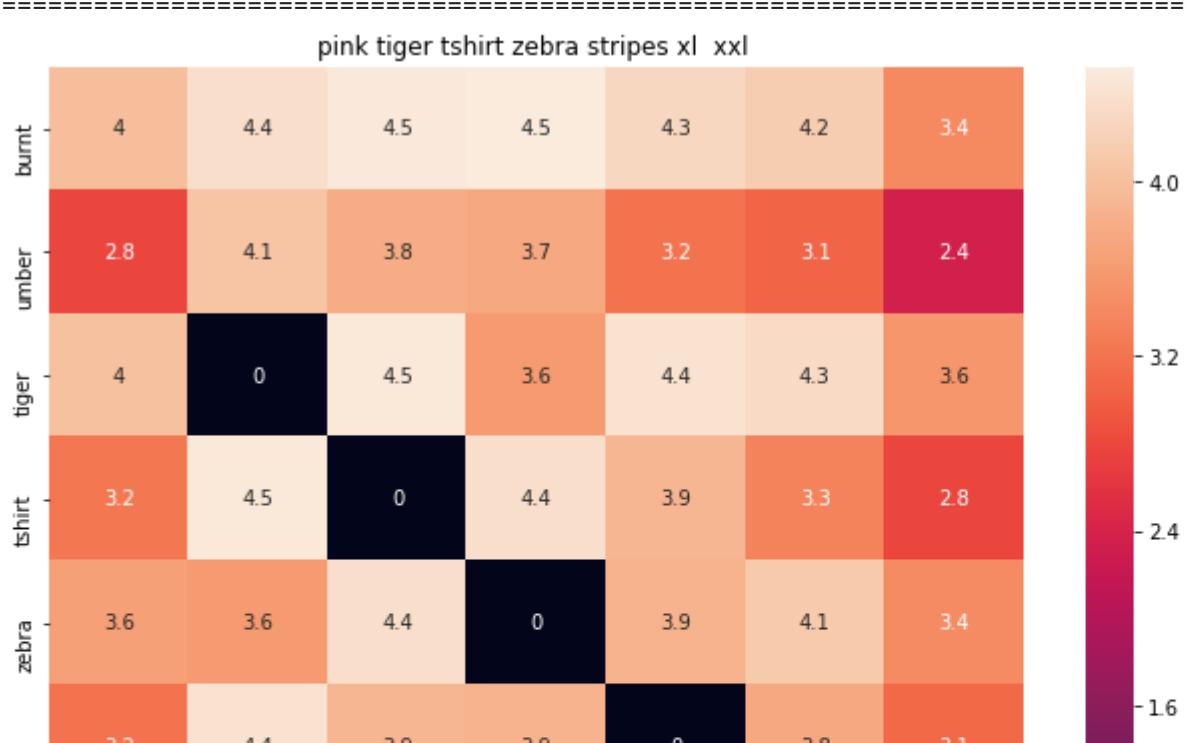
```
for i in range(0, len(indices)):
    heat_map_w2v(data['title'].loc[df_indices[0]], data['title'].loc[df_indices[i]], data[
        print('ASIN :', data['asin'].loc[df_indices[i]])]
        print('BRAND :', data['brand'].loc[df_indices[i]])]
        print('euclidean distance from given input image :', pdists[i])
        print('*125)
```

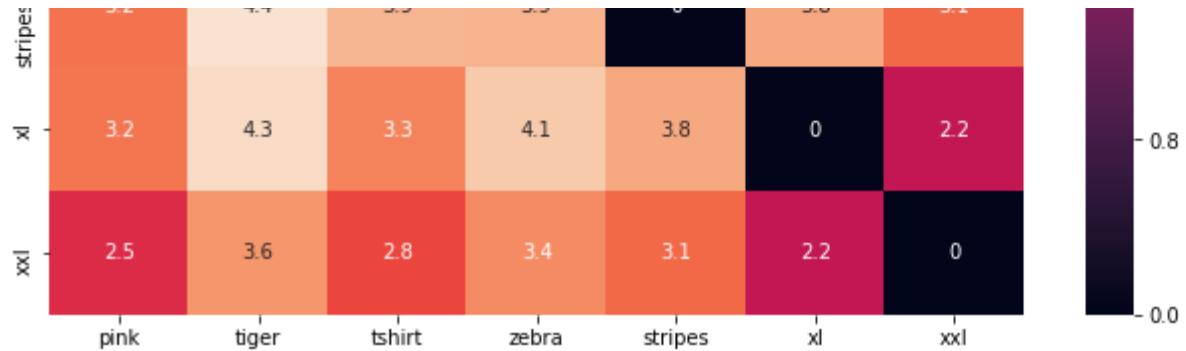
```
avg_w2v_model(12566, 20)
# in the give heat map, each cell contains the euclidean distance between words i, j
```





Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#). X

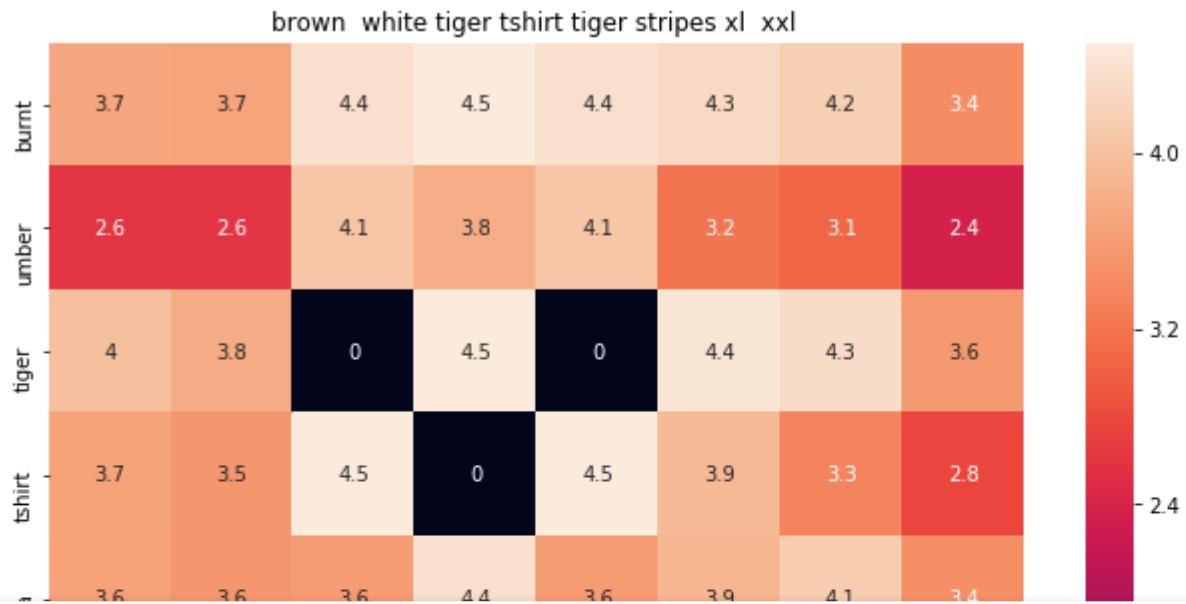




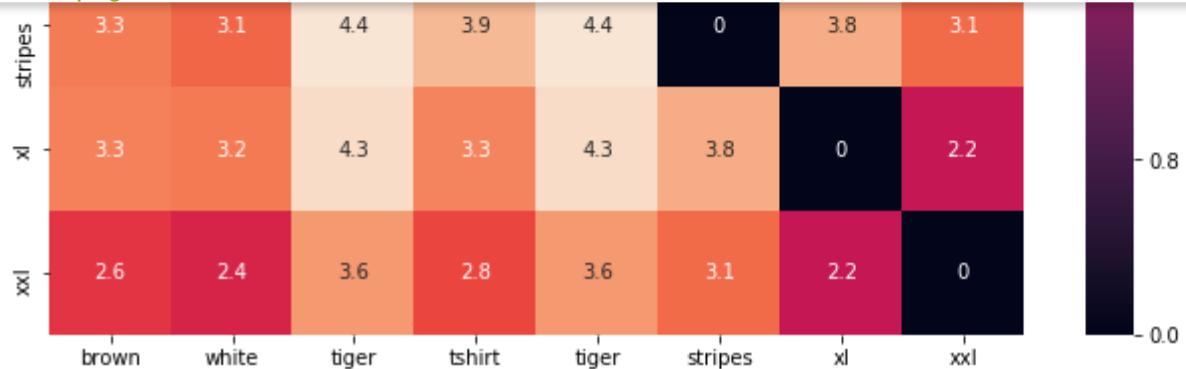
ASIN : B00JXQASS6

BRAND : Si Row

euclidean distance from given input image : 0.5891928



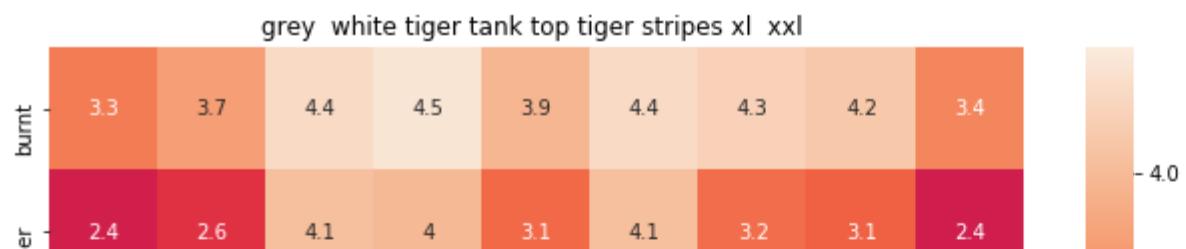
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

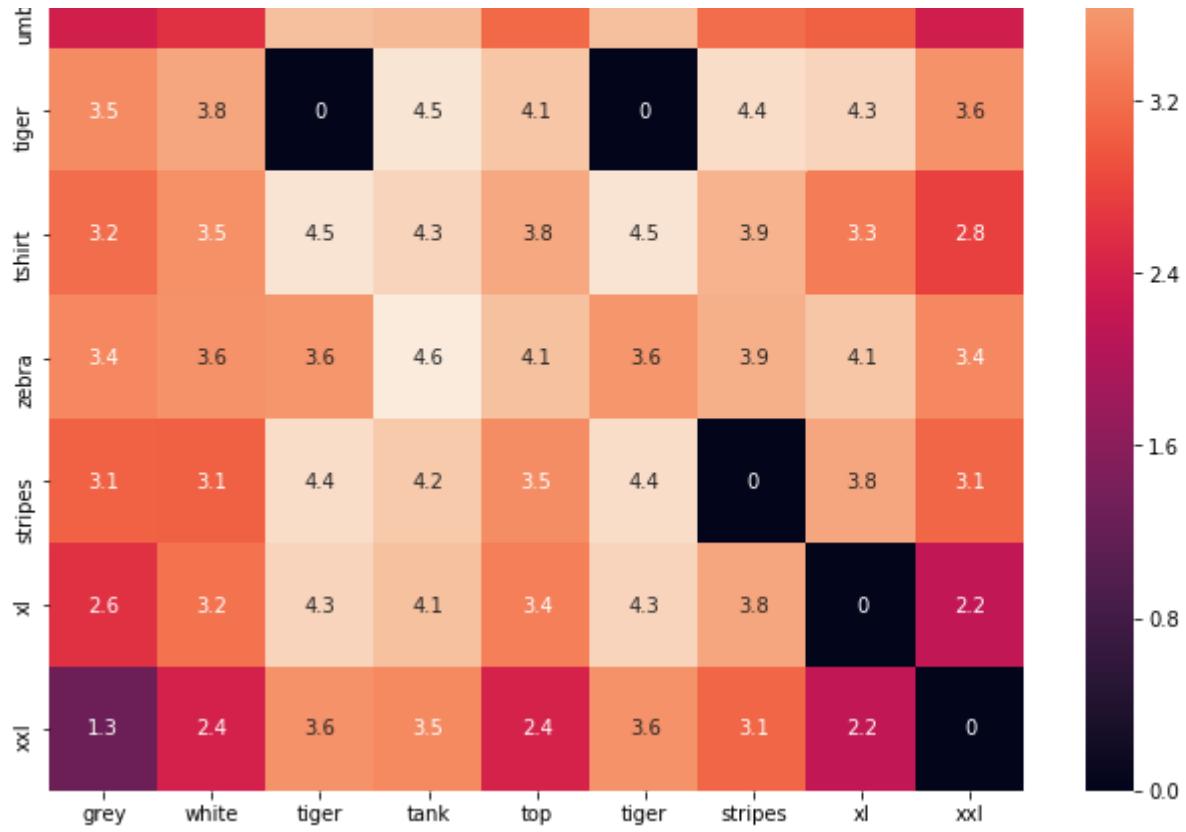


ASIN : B00JXQCWT0

BRAND : Si Row

euclidean distance from given input image : 0.7003436

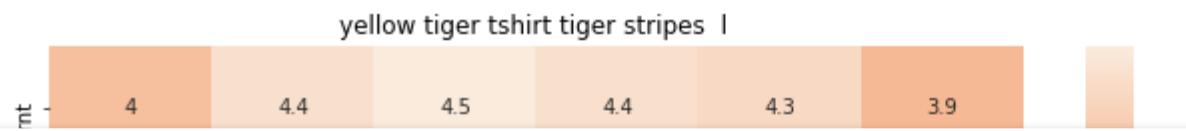




ASIN : B00JXQAFZ2

BRAND : Si Row

euclidean distance from given input image : 0.8928398



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

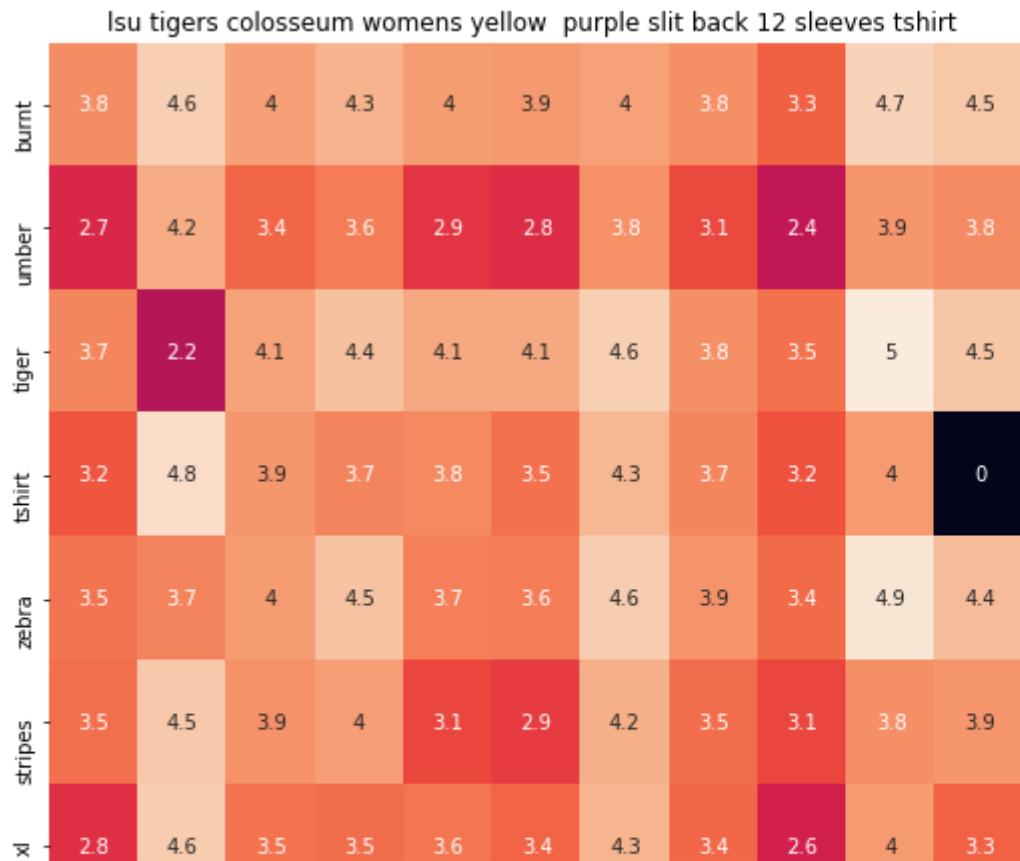




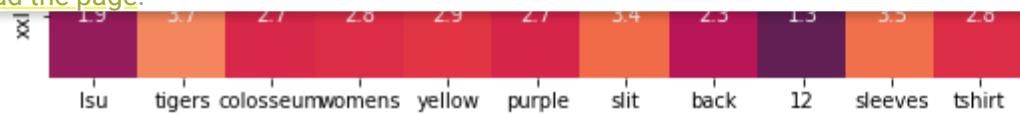
ASIN : B00JXQCUIC

BRAND : Si Row

euclidean distance from given input image : 0.9560129



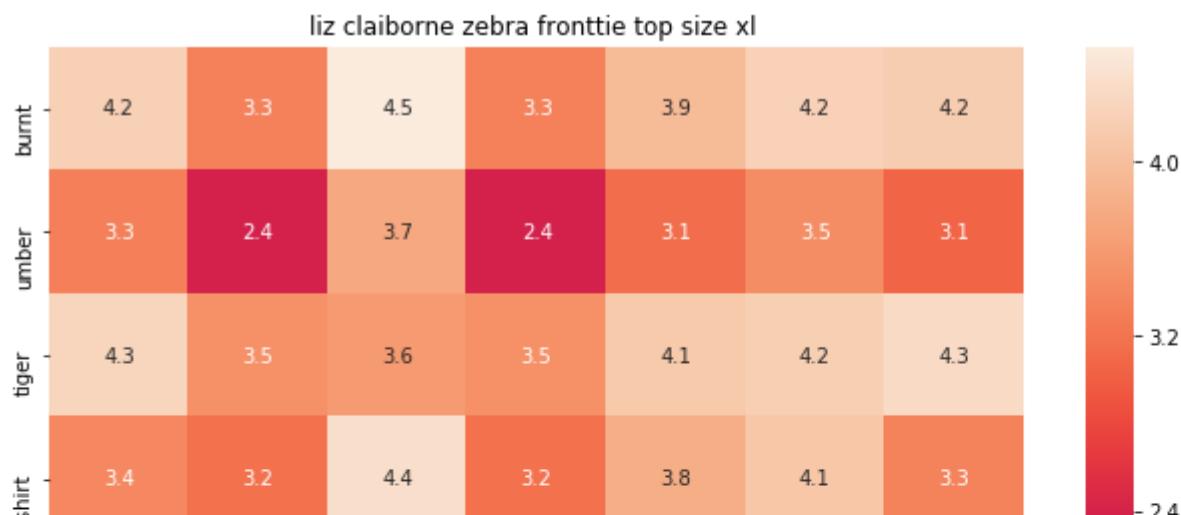
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

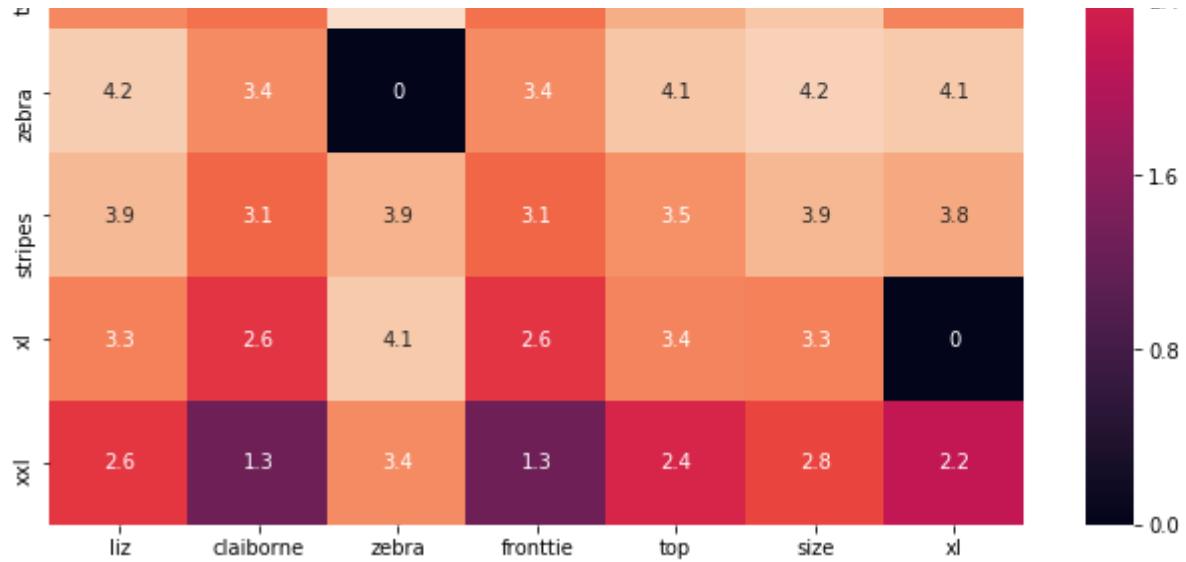


ASIN : B073R5Q8HD

BRAND : Colosseum

euclidean distance from given input image : 1.0229691





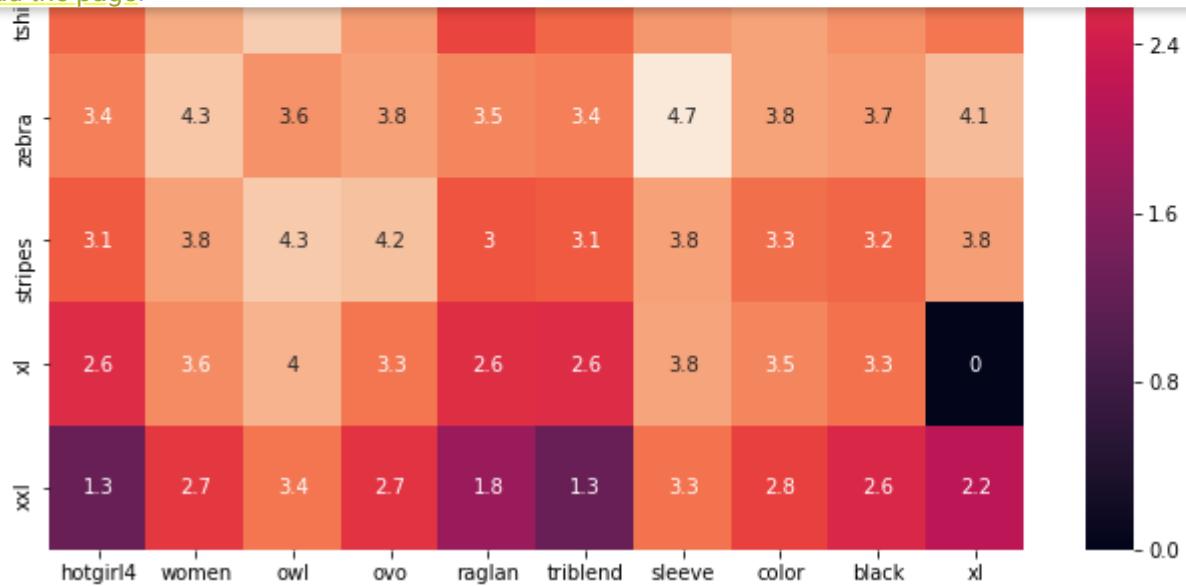
ASIN : B06XBY5QXL

BRAND : Liz Claiborne

euclidean distance from given input image : 1.0669324



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B01L8L73M2

BRAND : Hotgirl4 Raglan Design

euclidean distance from given input image : 1.0731406

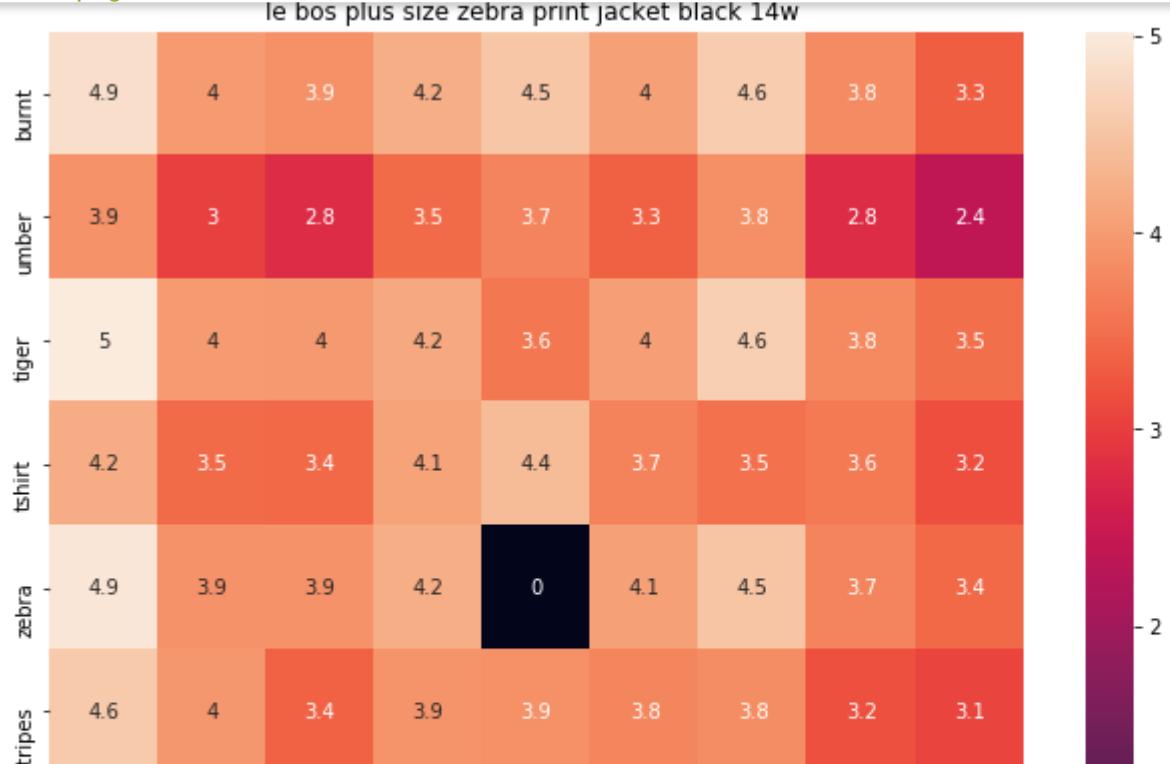
===== vansty hippo naughty round neck tshirt women yellow size xl

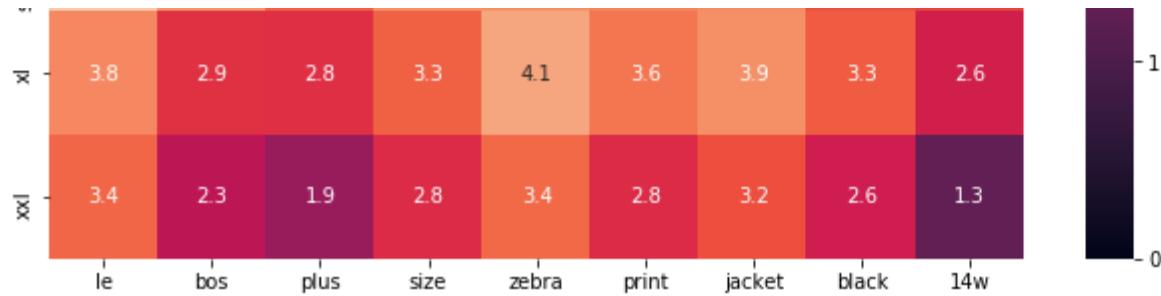


ASIN : B01EJS5H06

BRAND : Vansty

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X





ASIN : B01B01XRK8

BRAND : Le Bos

euclidean distance from given input image : 1.0839965



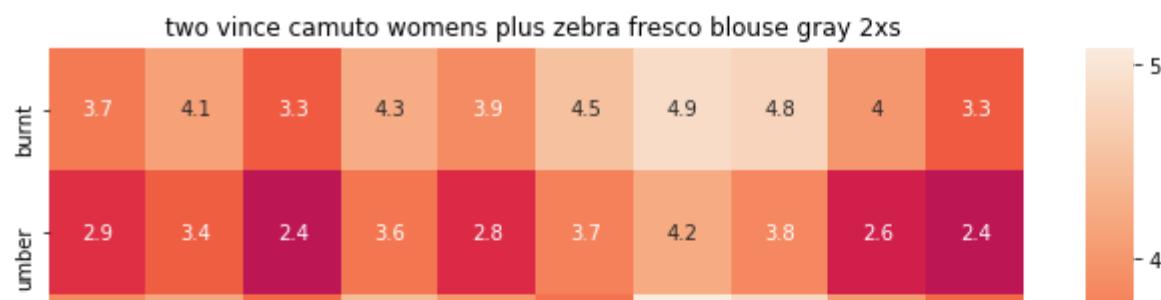
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

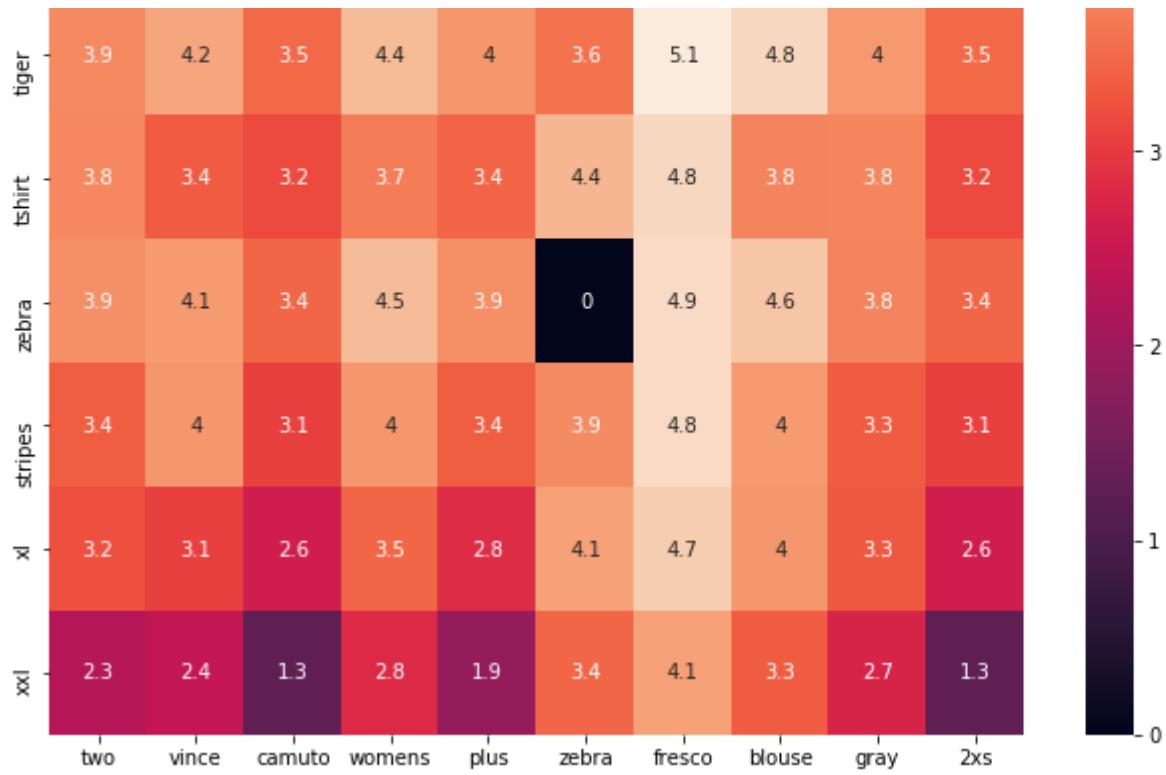


ASIN : B072R2JXKW

BRAND : WHAT ON EARTH

euclidean distance from given input image : 1.0842218

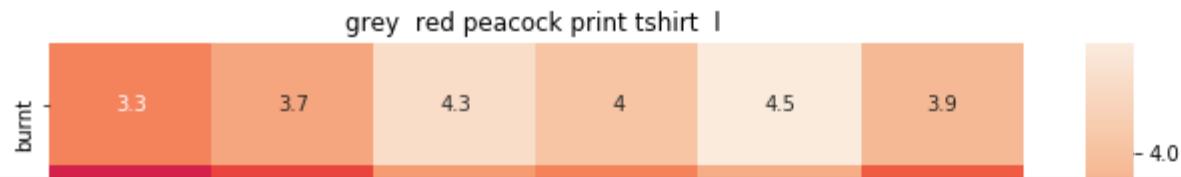




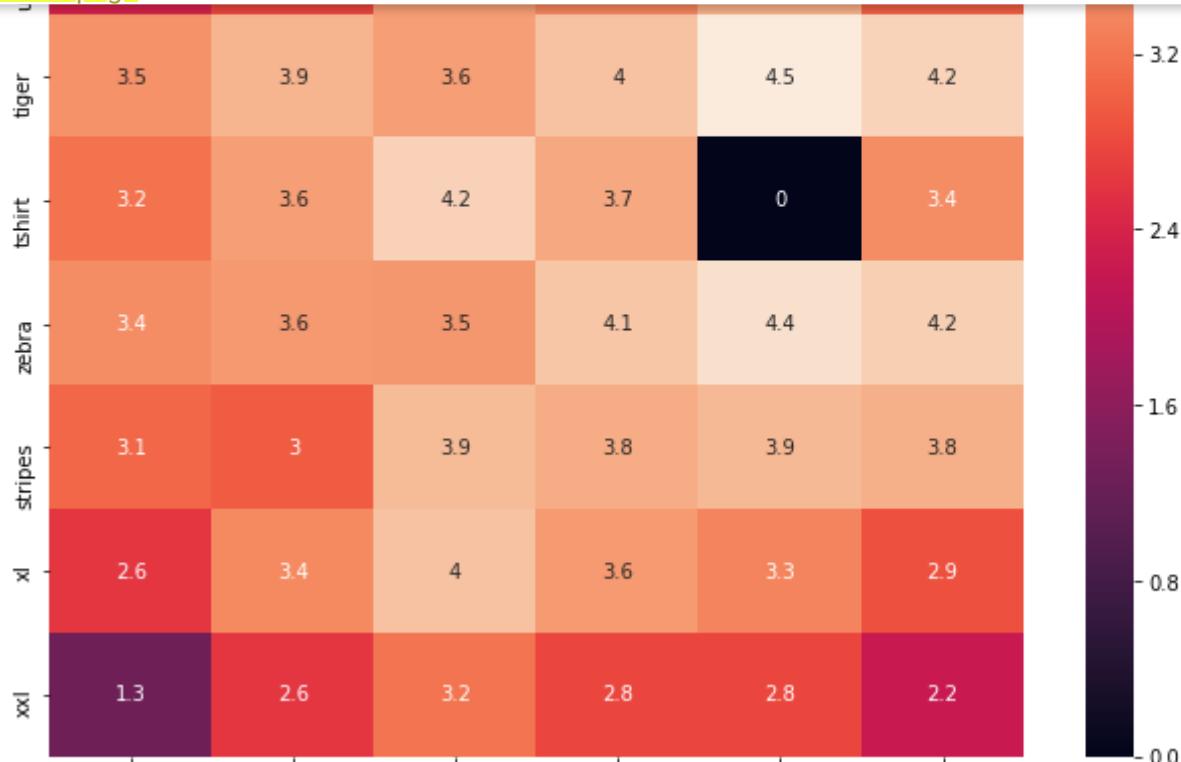
ASIN : B074MJRGW6

BRAND : Two by Vince Camuto

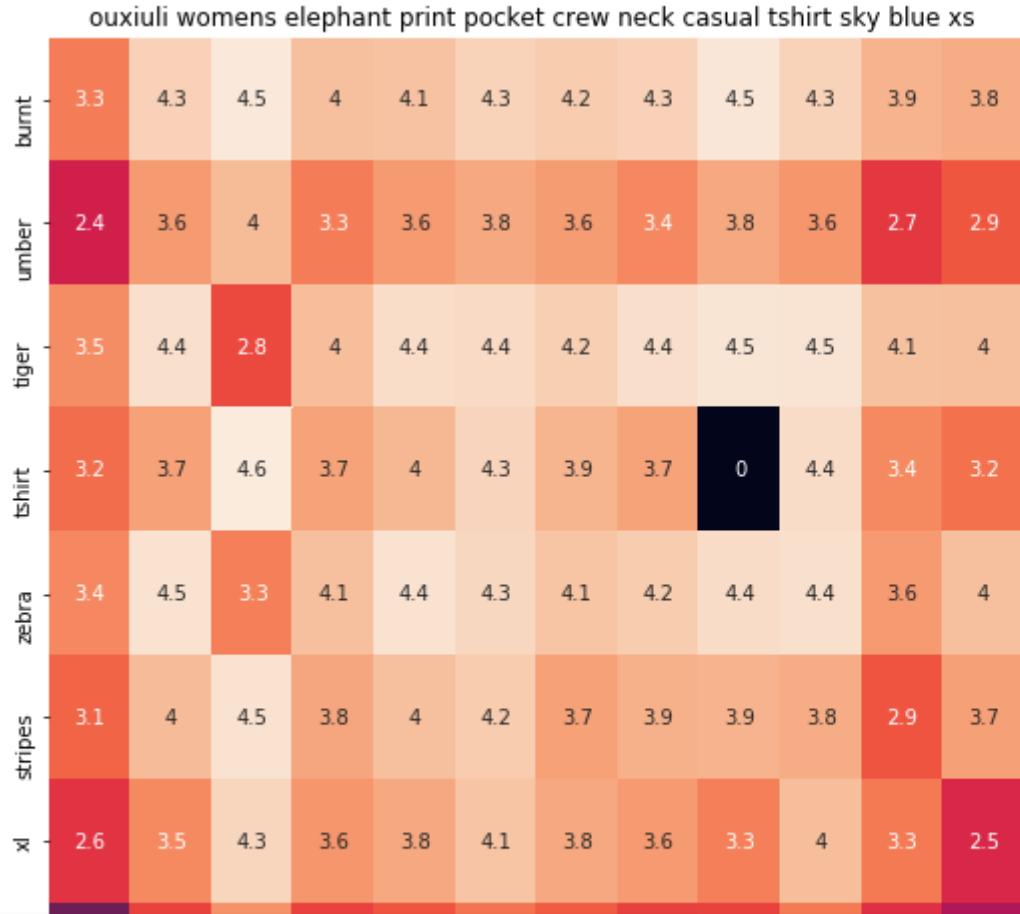
euclidean distance from given input image : 1.0895039



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

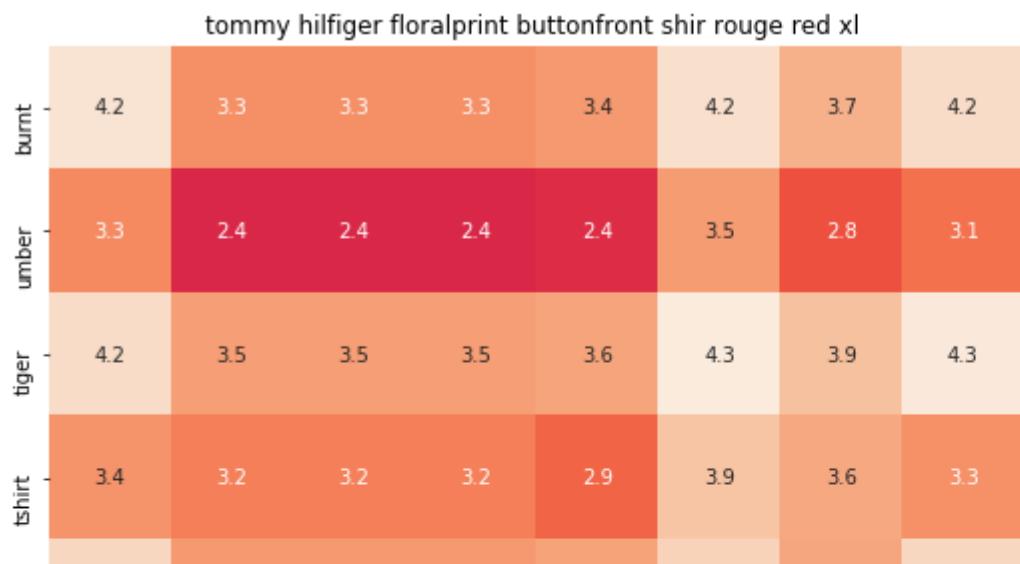


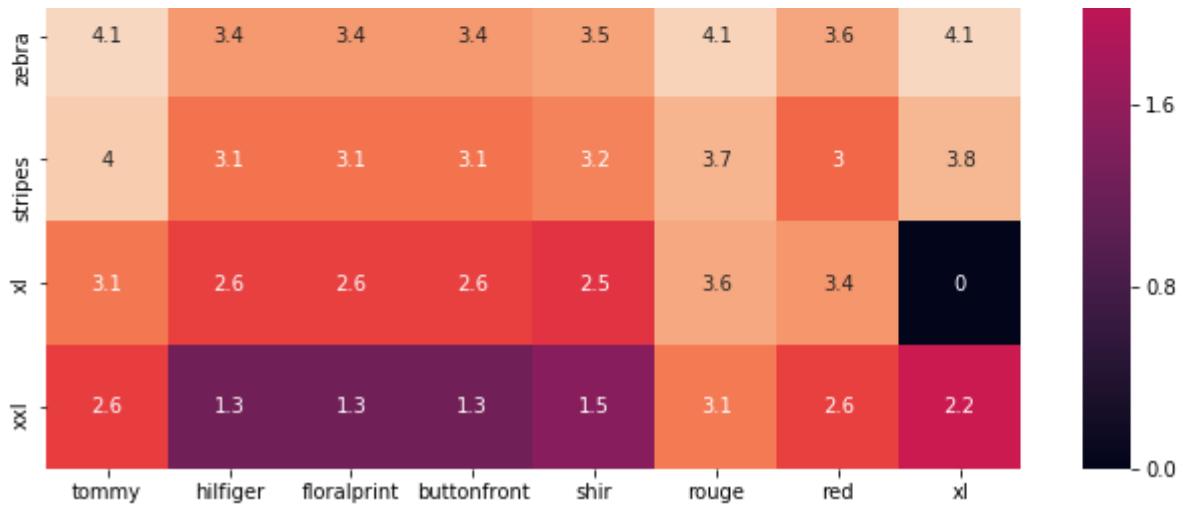
ASIN : B00JXQCFRS
 BRAND : Si Row
 euclidean distance from given input image : 1.0900588



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

ASIN : B01I53HU6K
 BRAND : ouxiuli
 euclidean distance from given input image : 1.0920112

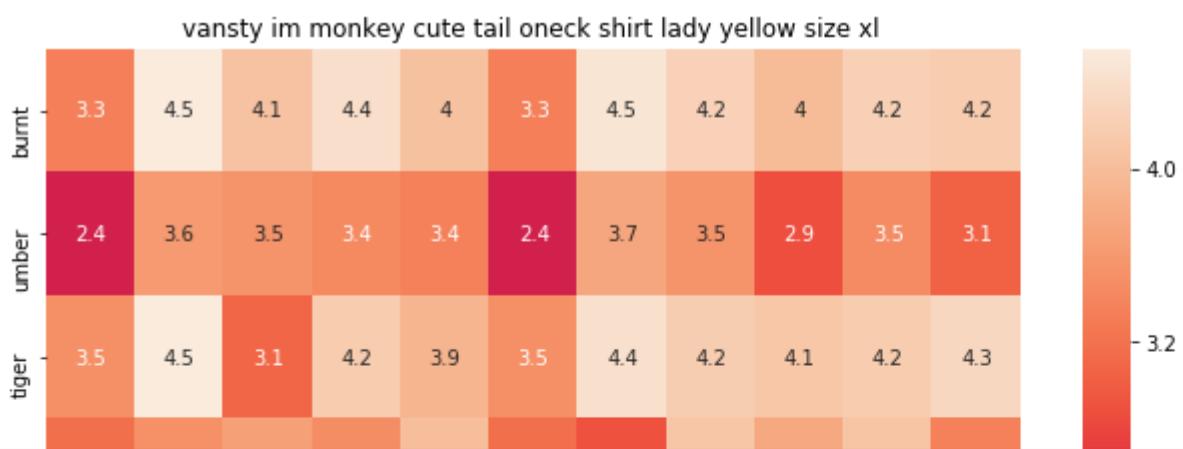




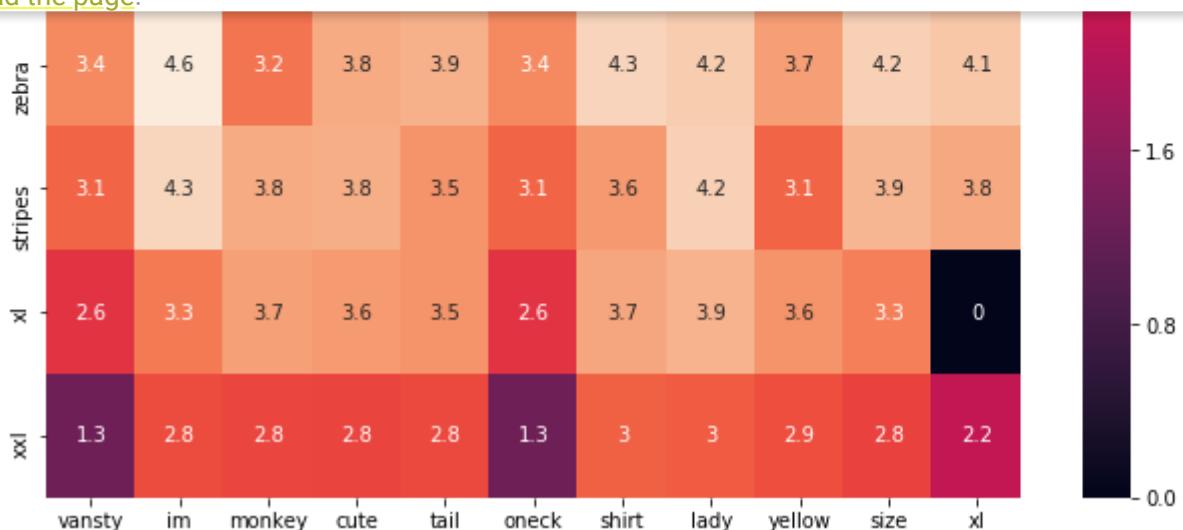
ASIN : B0711NGTQM

BRAND : THILFIGER RTW

euclidean distance from given input image : 1.0923418



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

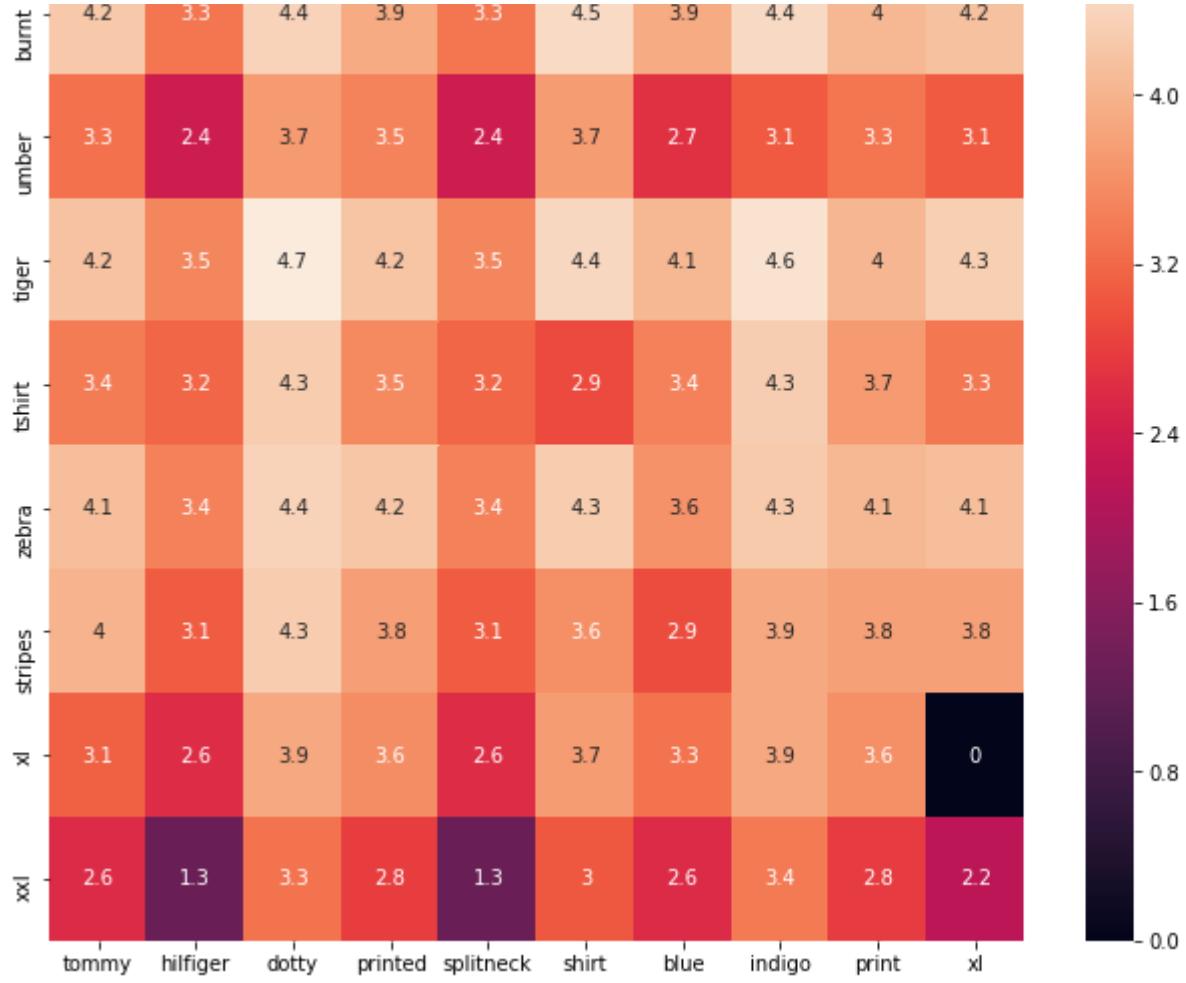


ASIN : B01EFSL08Y

BRAND : Vansty

euclidean distance from given input image : 1.0934006



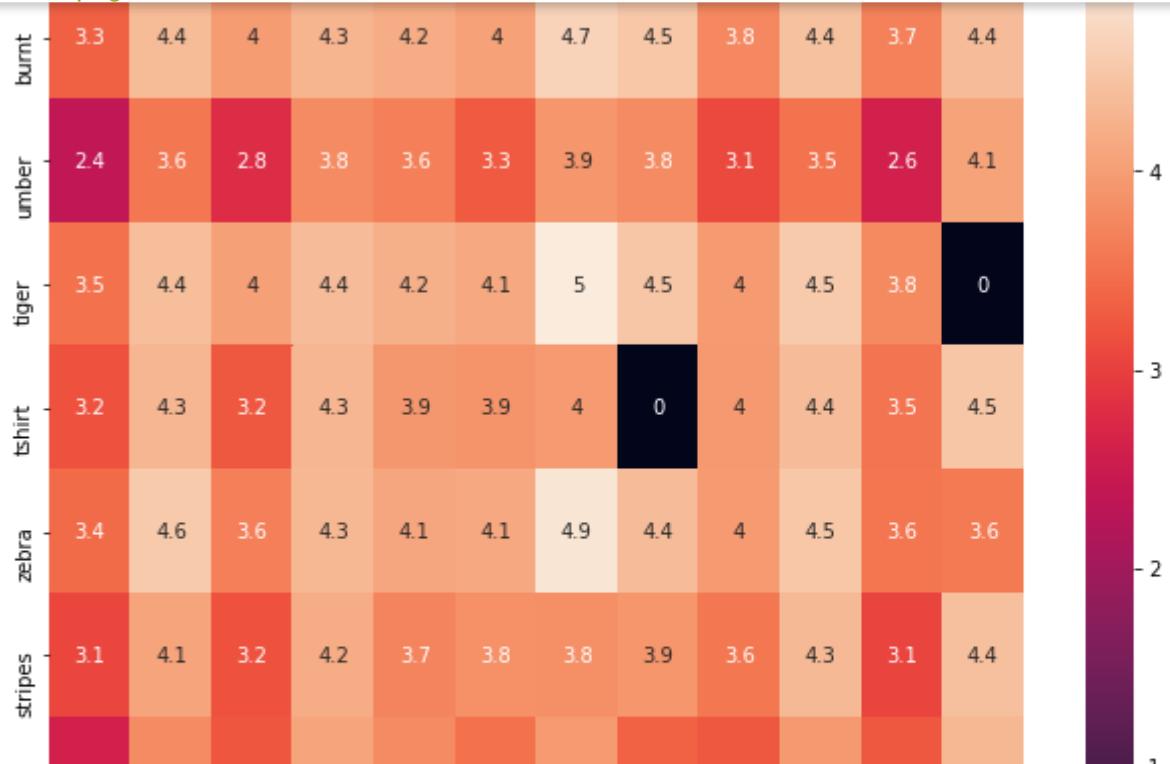


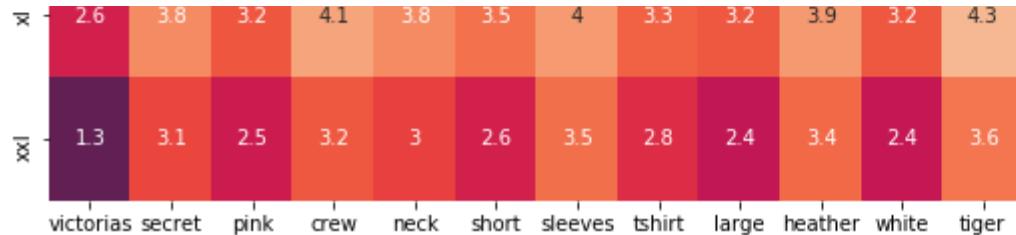
ASIN : B0716TVWQ4

BRAND : THILFIGER RTW

euclidean distance from given input image : 1.0942025

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

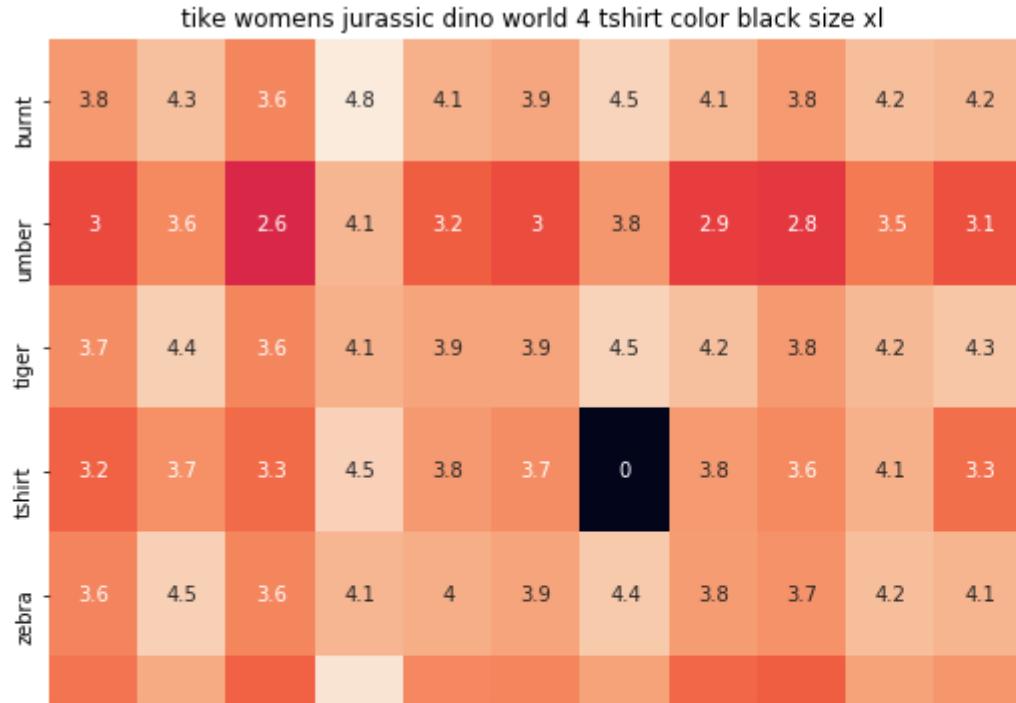




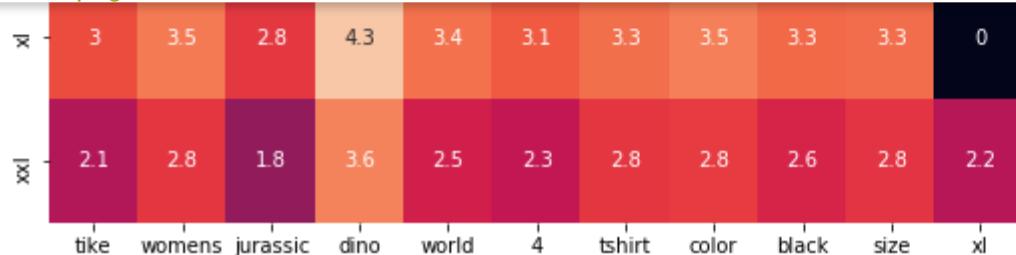
ASIN : B0716MVPGV

BRAND : V.Secret

euclidean distance from given input image : 1.0948305



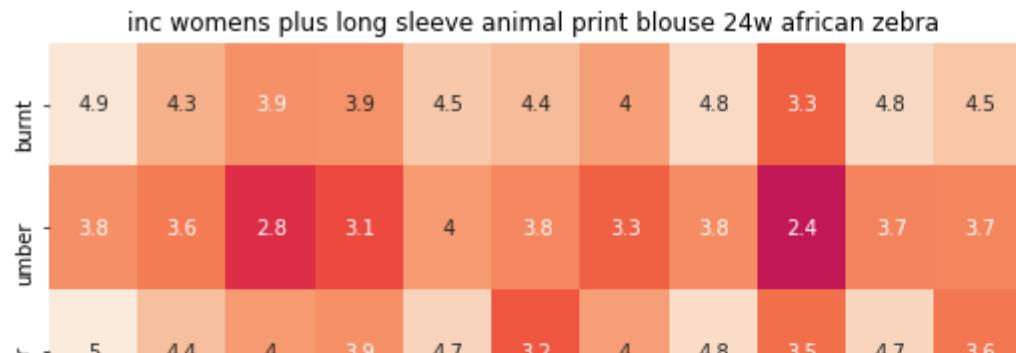
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

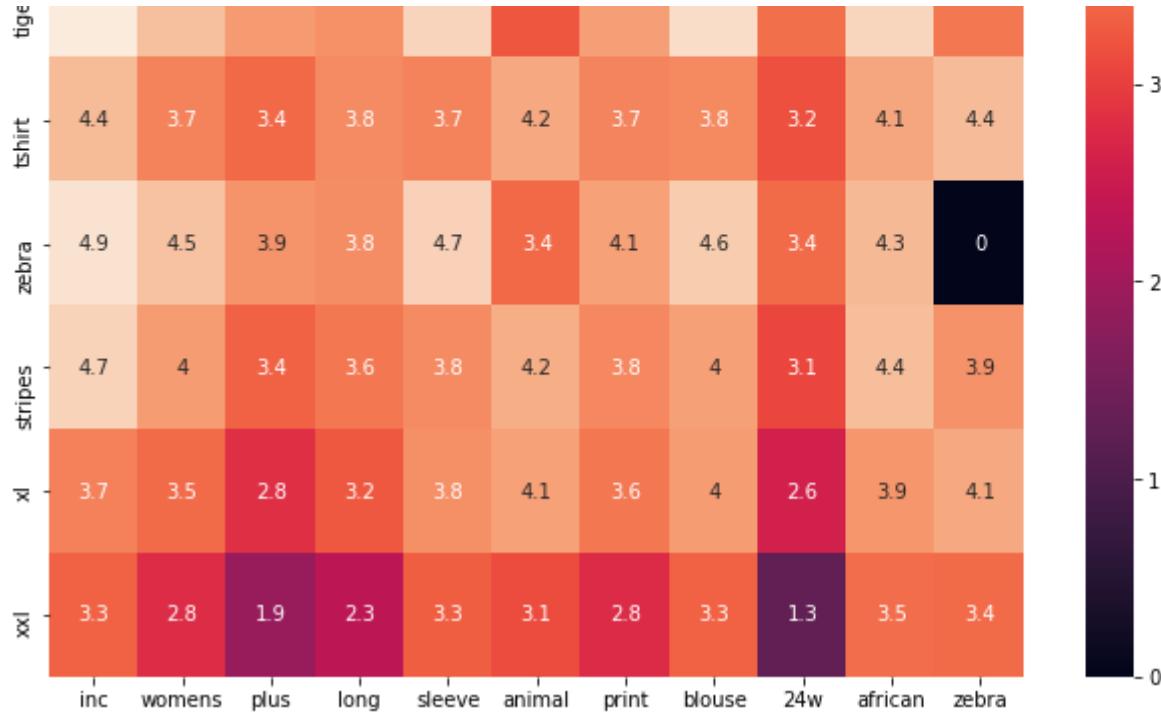


ASIN : B0160PN40I

BRAND : TIKE Fashions

euclidean distance from given input image : 1.0951276





ASIN : B018WDJCUA

BRAND : INC - International Concepts Woman

euclidean distance from given input image : 1.0966893

=====

▼ [9.4] IDF weighted Word2Vec for product similarity

doc_id = 0

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```

for i in data['title']:
    w2v_title_weight.append(build_avg_vec(i, 300, doc_id,'weighted'))
    doc_id += 1
# w2v_title = np.array(# number of doc in courpus * 300), each row corresponds to a doc
w2v_title_weight = np.array(w2v_title_weight)

def weighted_w2v_model(doc_id, num_results):
    # doc_id: apparel's id in given corpus

    # pairwise_dist will store the distance from given input apparel to all remaining apparel
    # the metric we used here is cosine, the coside distance is mesured as K(X, Y) = <X, Y> /
    # http://scikit-learn.org/stable/modules/metrics.html#cosine-similarity
    pairwise_dist = pairwise_distances(w2v_title_weight, w2v_title_weight[doc_id].reshape(1,-1))

    # np.argsort will return indices of 9 smallest distances
    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    #pdists will store the 9 smallest distances
    pdists = np.sort(pairwise_dist.flatten())[0:num_results]

    #data frame indices of the 9 smallest distace's

```

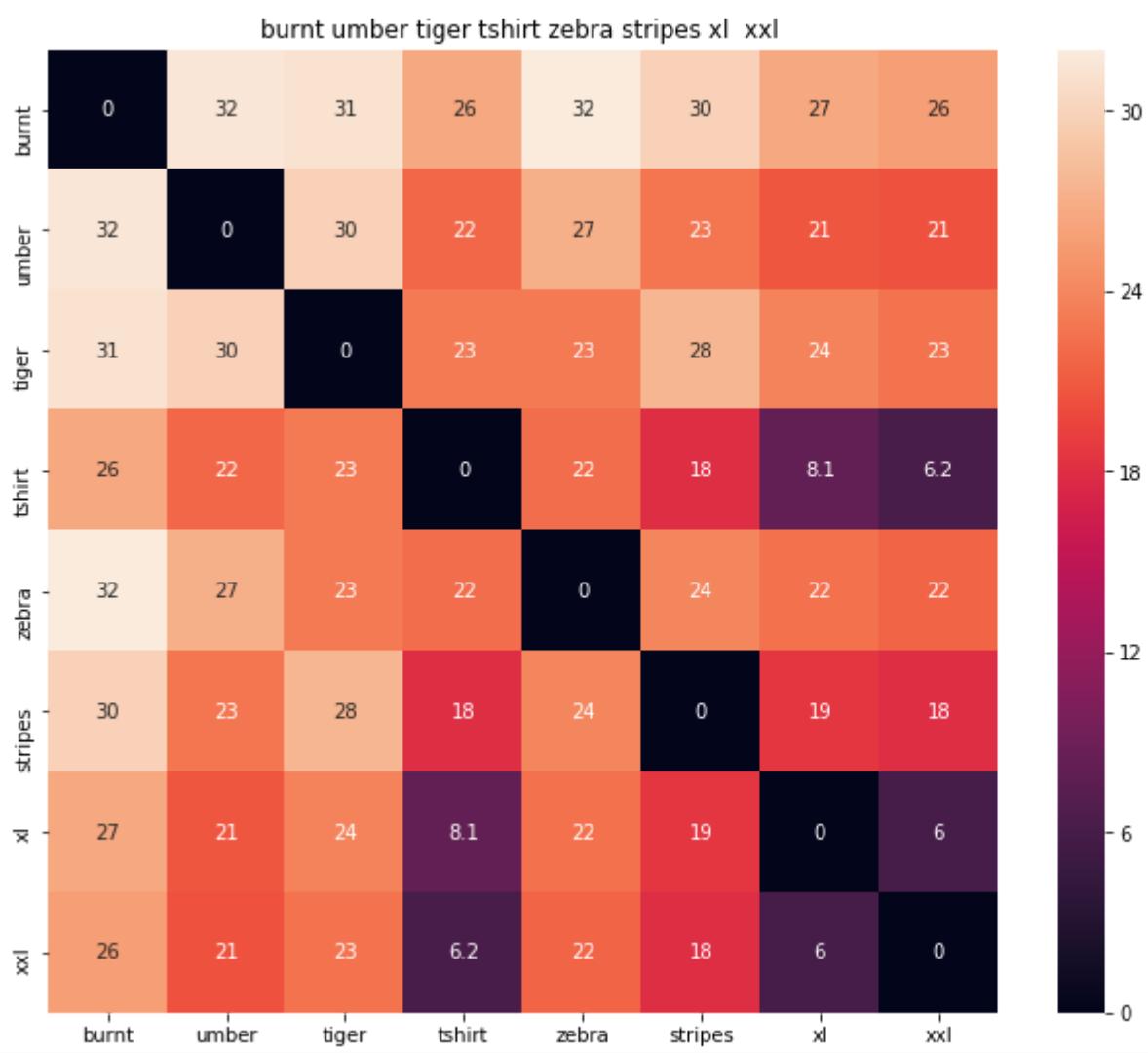
```
df_indices = list(data.index[indices])

for i in range(0, len(indices)):
    heat_map_w2v(data['title'].loc[df_indices[0]],data['title'].loc[df_indices[i]], data[
    print('ASIN :',data['asin'].loc[df_indices[i]])]
    print('Brand :',data['brand'].loc[df_indices[i]])]
    print('euclidean distance from input :', pdists[i])
    print('*'*125)

weighted_w2v_model(12566, 20)
#931
#12566
# in the give heat map, each cell contains the euclidean distance between words i, j
```

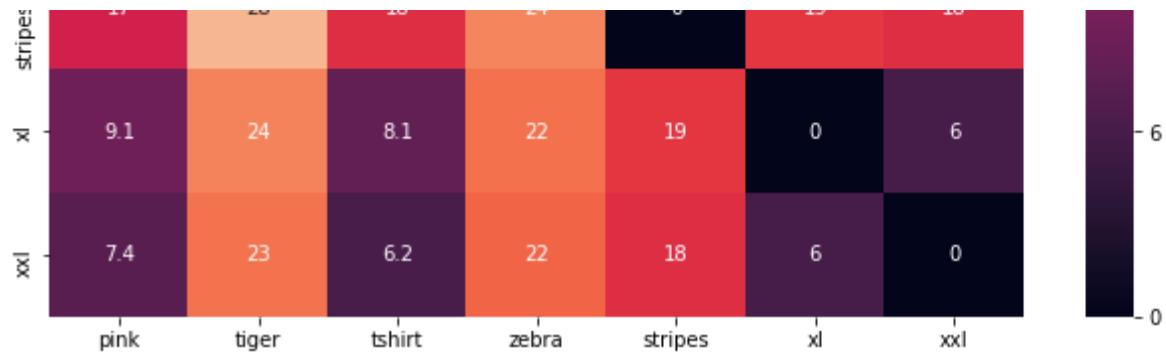


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#). X

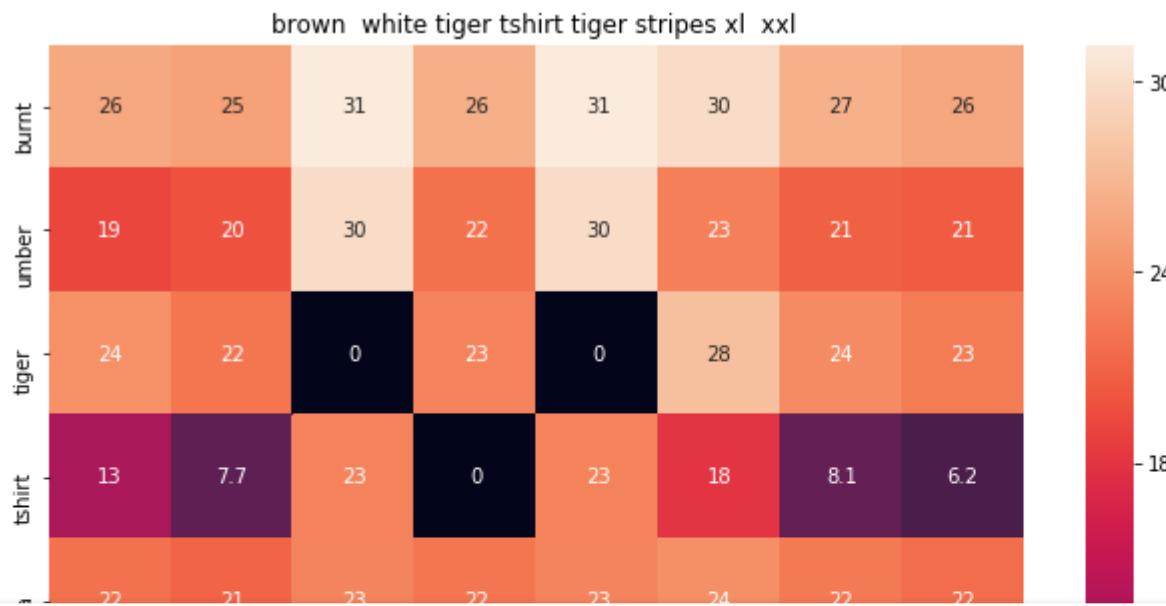




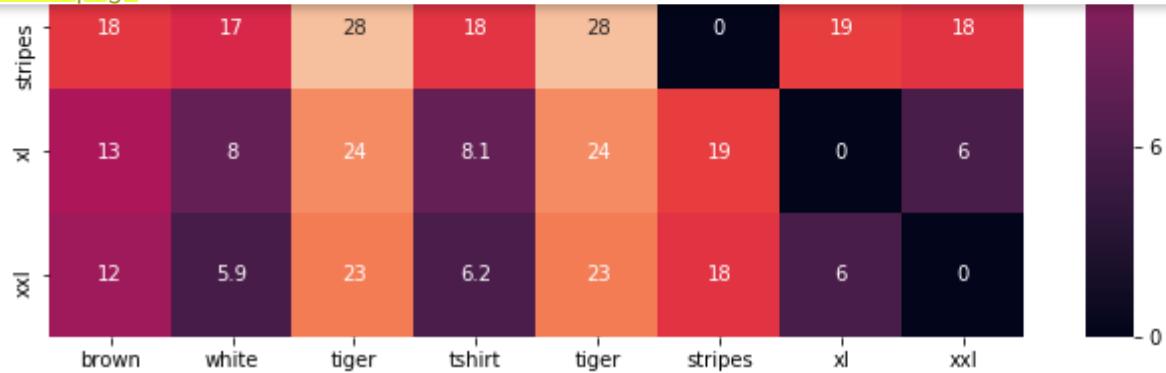
ASIN : B00JXQASS6

Brand : Si Row

euclidean distance from input : 4.0638876



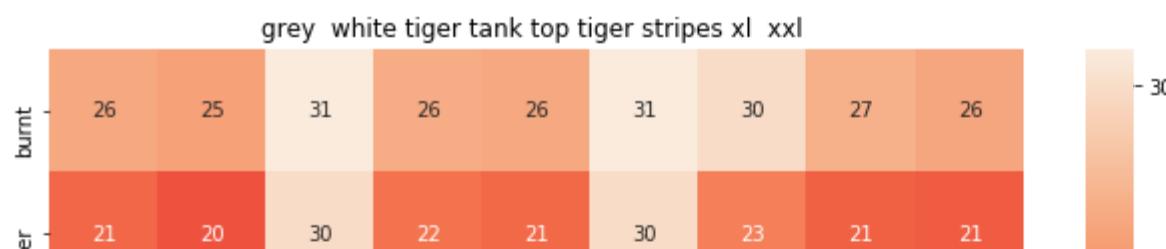
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

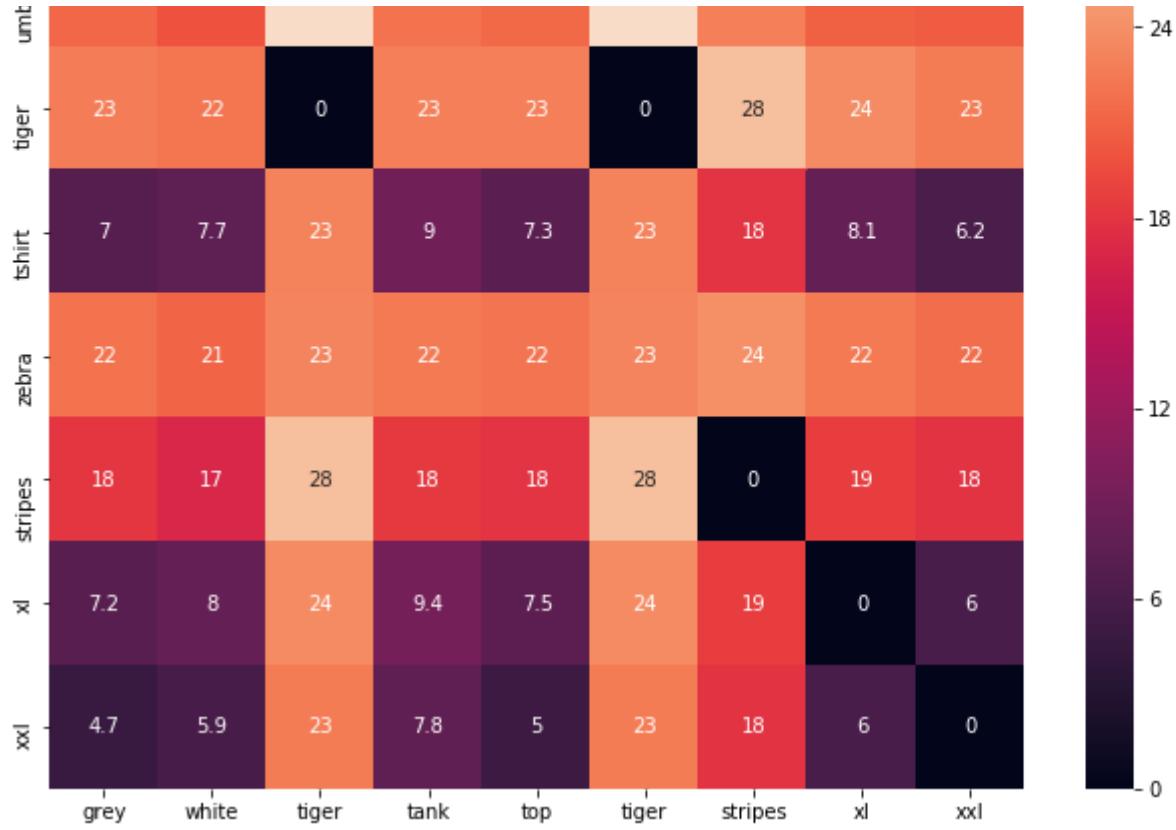


ASIN : B00JXQCWT0

Brand : Si Row

euclidean distance from input : 4.770942

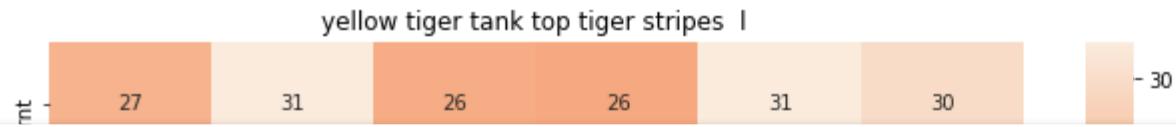




ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from input : 5.3601613



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

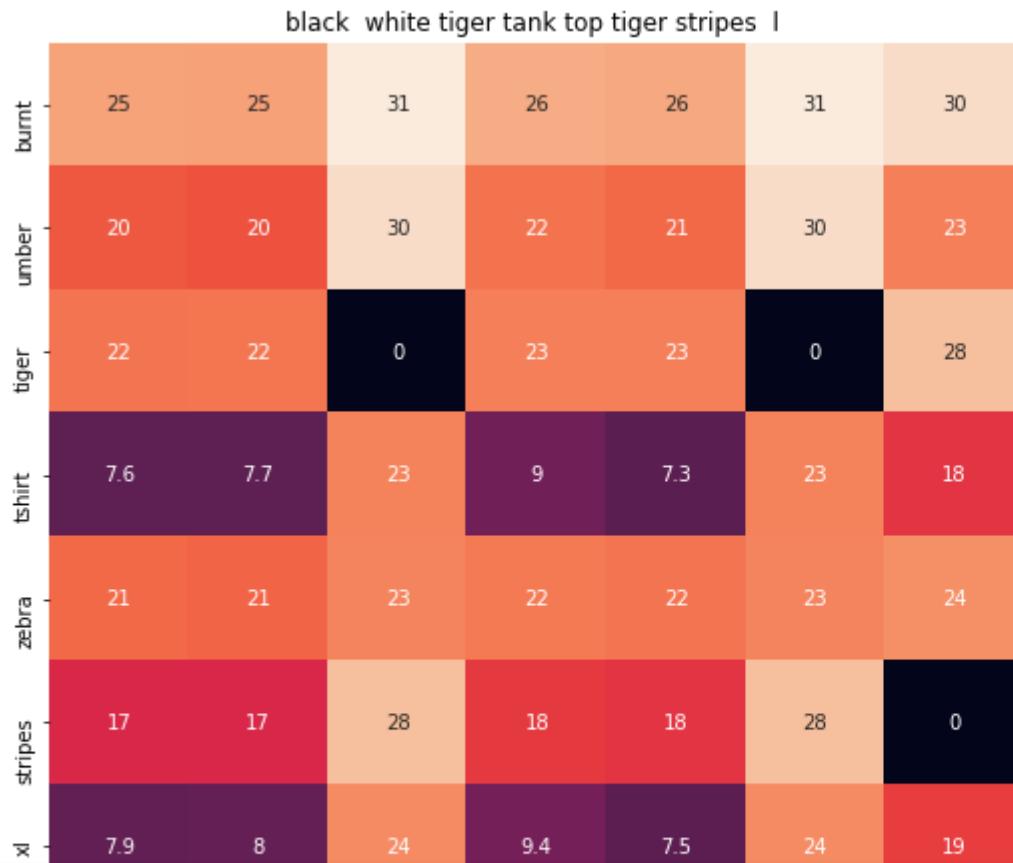




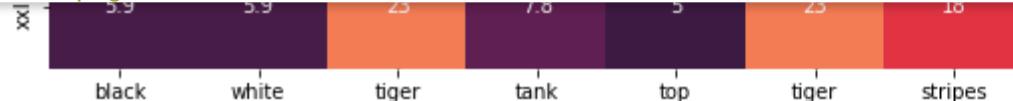
ASIN : B00JXQAUWA

Brand : Si Row

euclidean distance from input : 5.689523



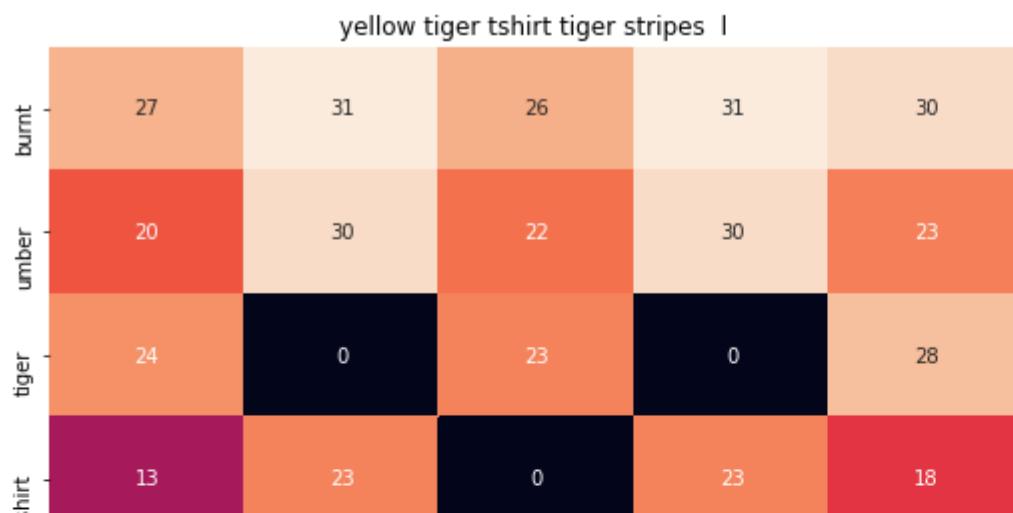
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

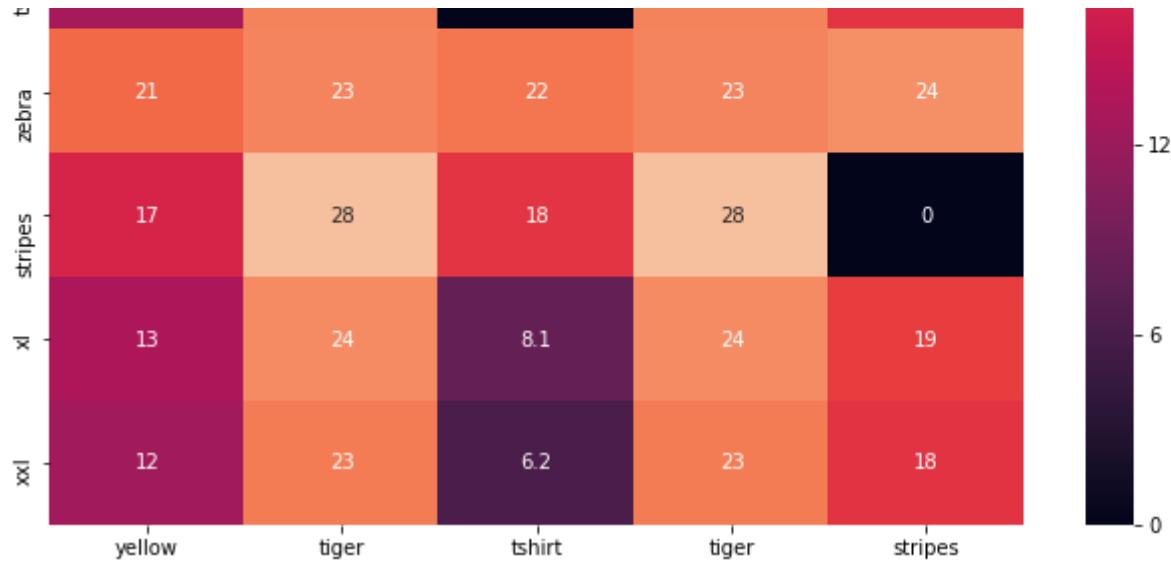


ASIN : B00JXQA094

Brand : Si Row

euclidean distance from input : 5.6930227

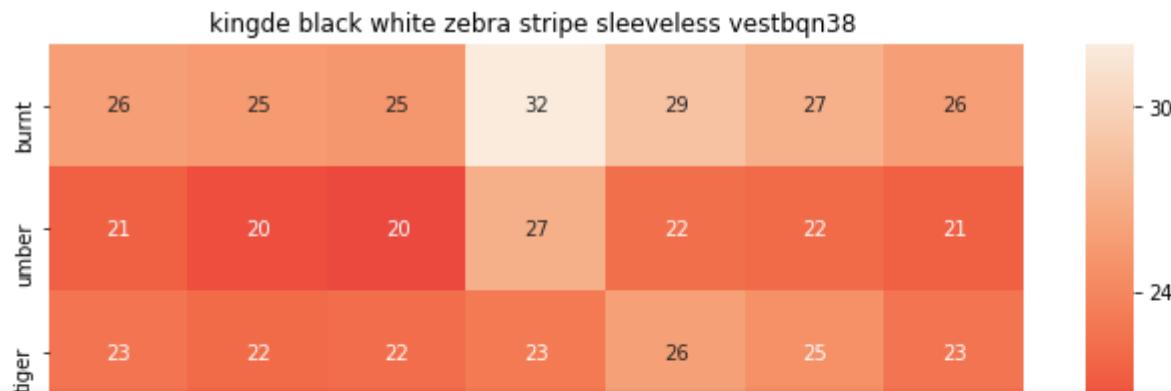




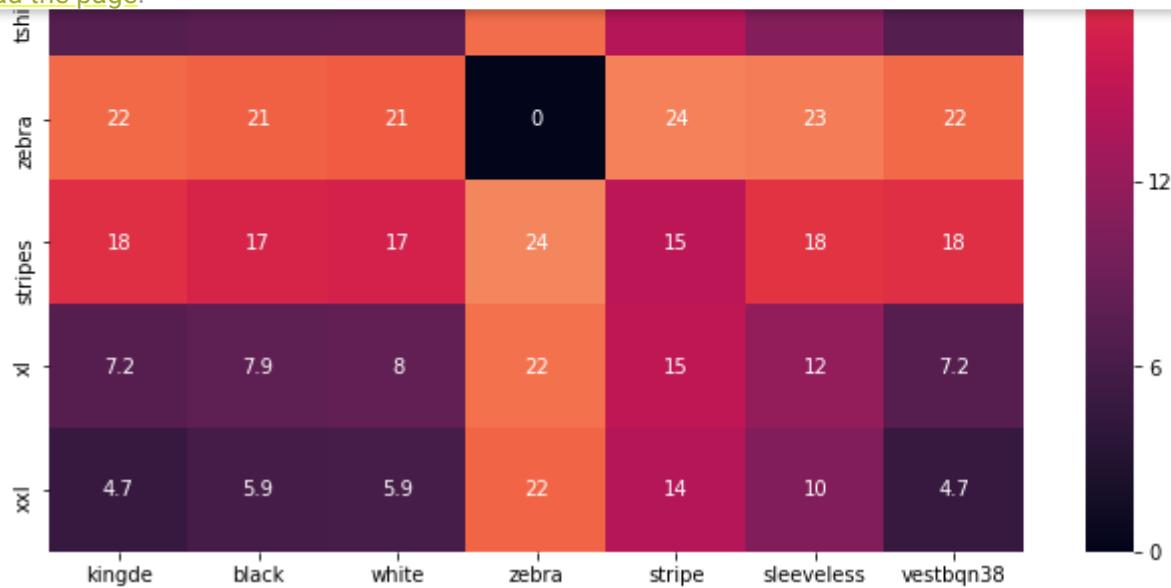
ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from input : 5.8934426



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

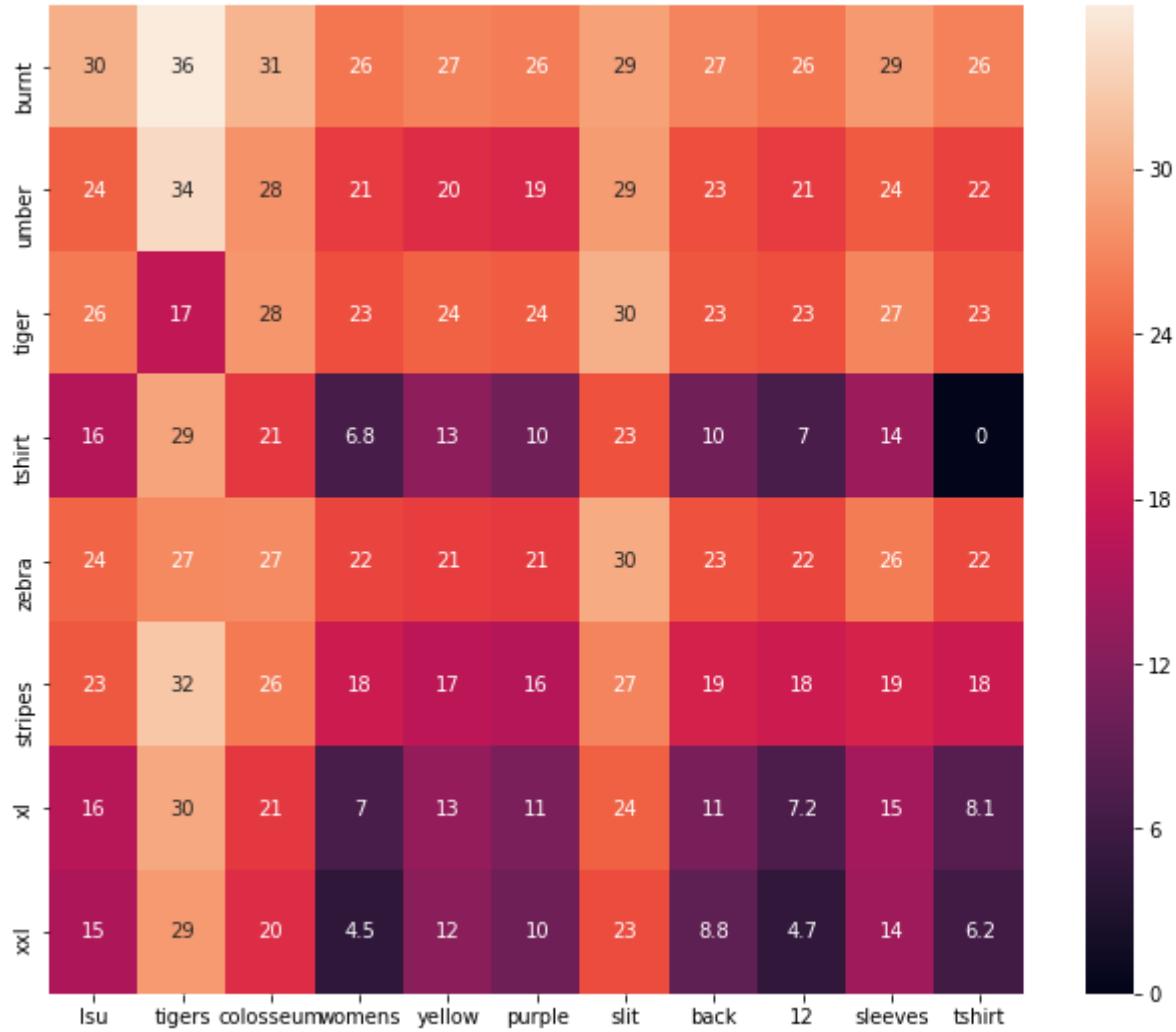


ASIN : B015H41F6G

Brand : KINGDE

euclidean distance from input : 6.13299

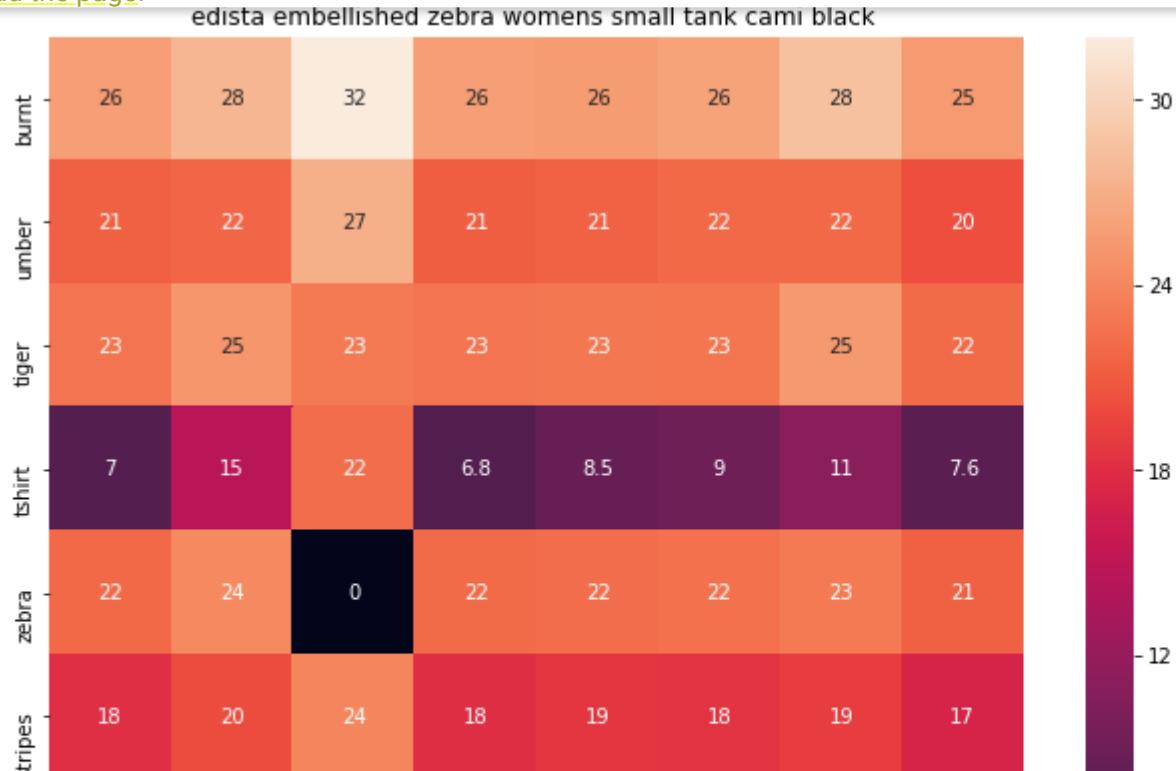
Isu tigers colosseum womens yellow purple slit back 12 sleeves tshirt

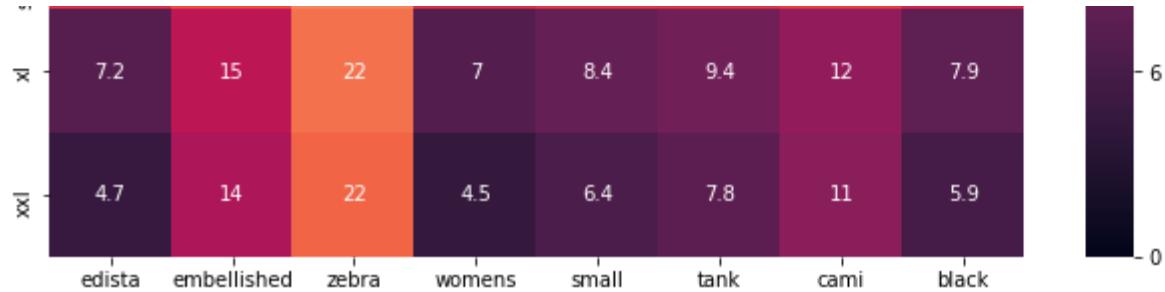


ASIN : B073R5Q8HD

Brand : Colosseum

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

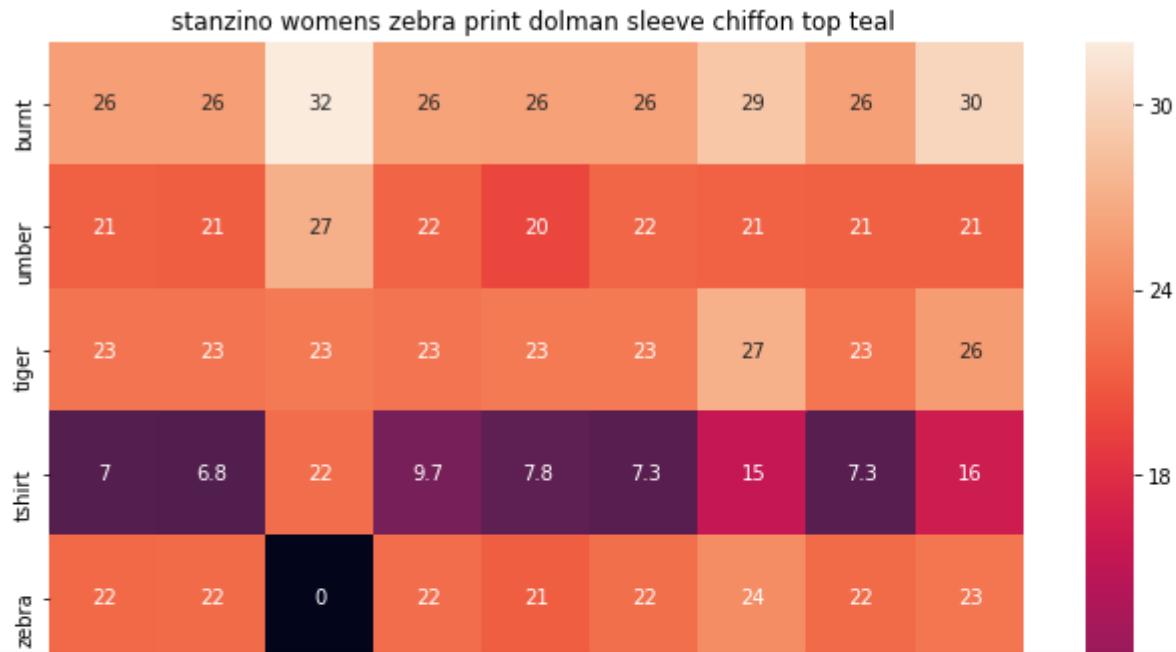




ASIN : B074P8MD22

Brand : Edista

euclidean distance from input : 6.3922043



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

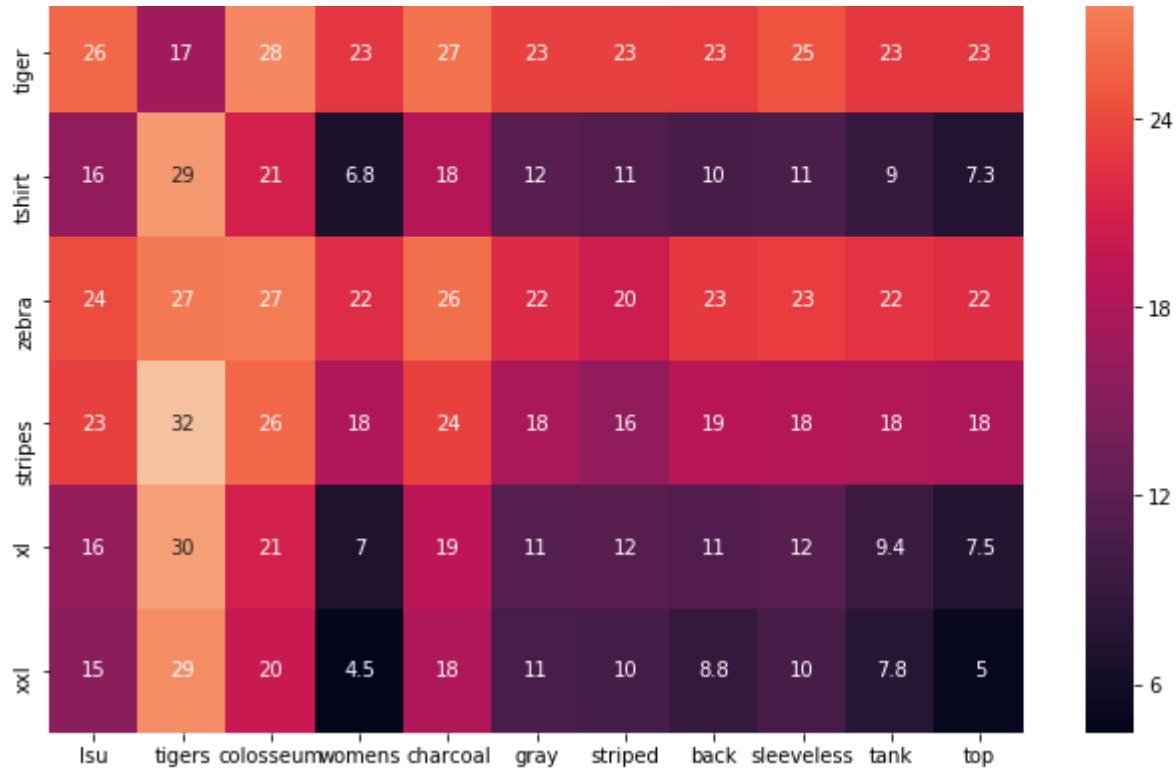


ASIN : B00C0I3U3E

Brand : Stanzino

euclidean distance from input : 6.414901





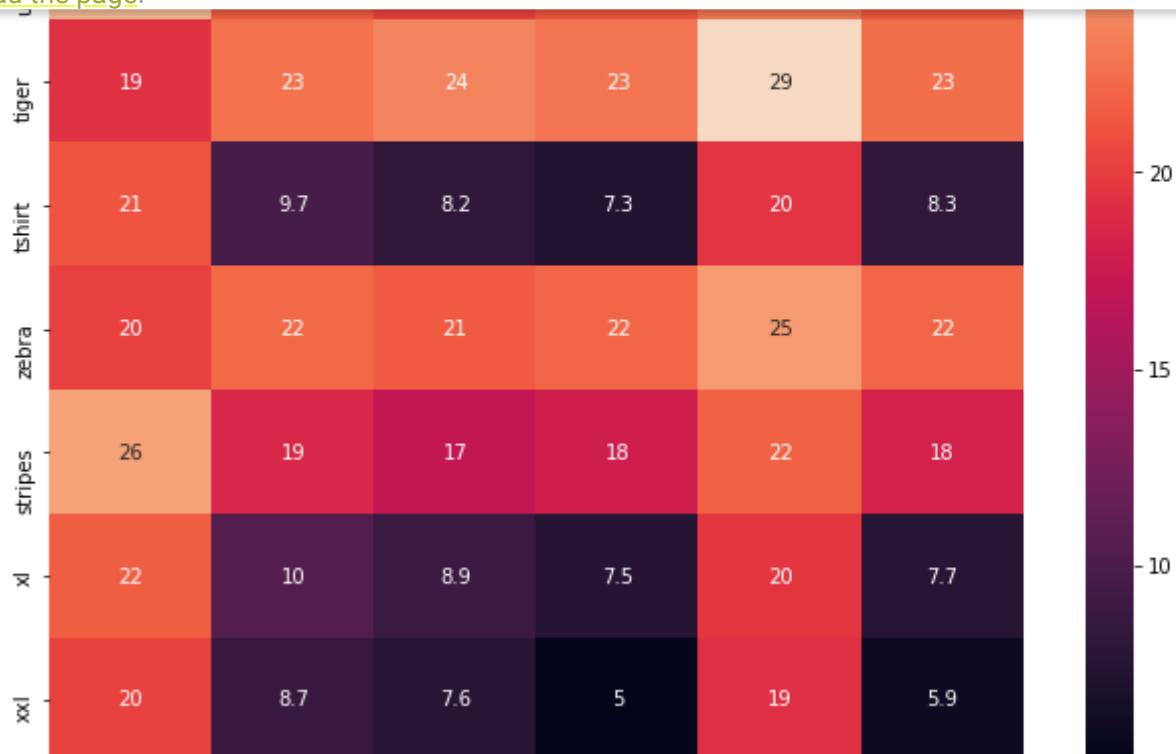
ASIN : B073R4ZM7Y

Brand : Colosseum

euclidean distance from input : 6.4509606



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

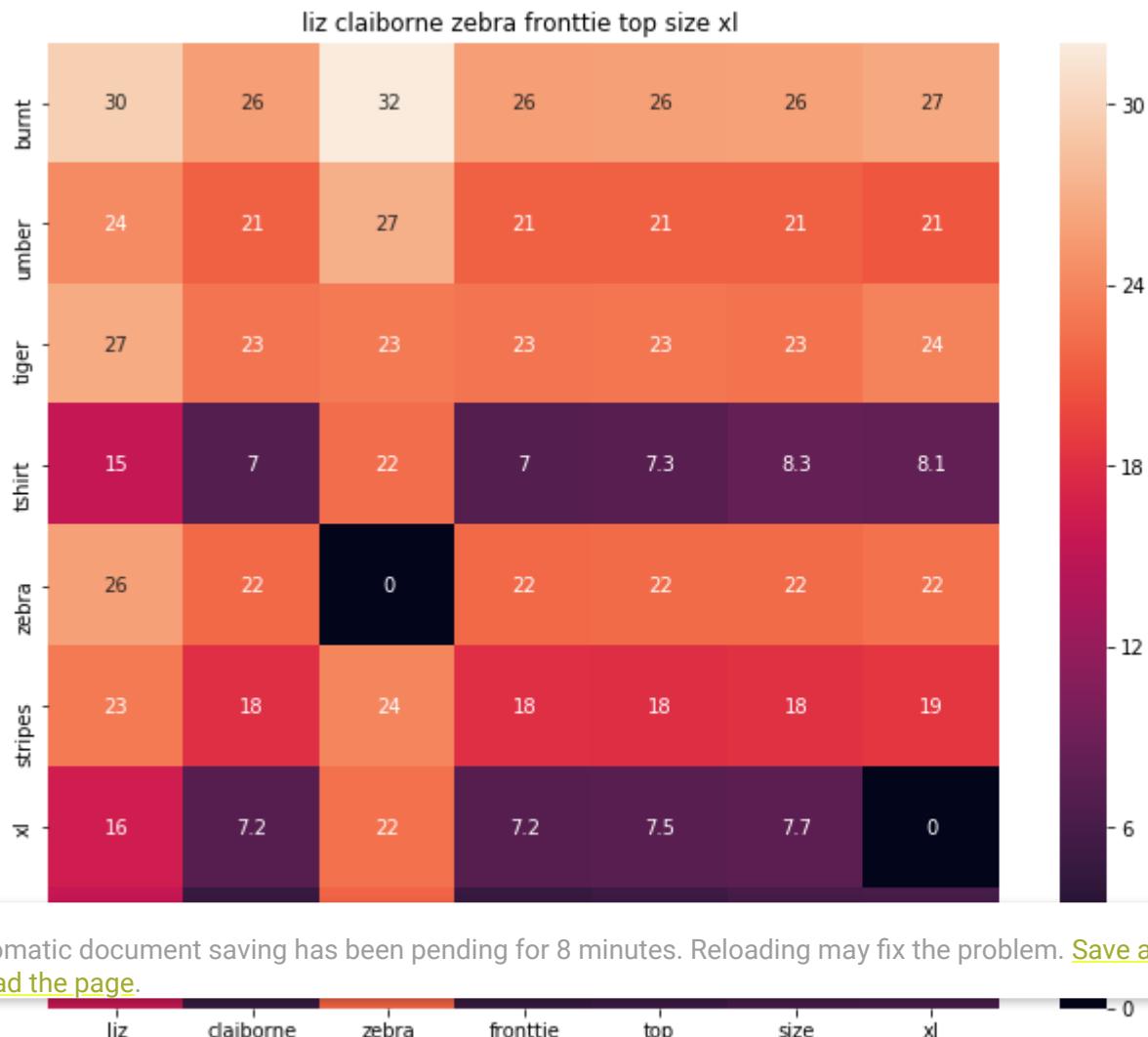


leopard print raglan top burgundy size

ASIN : B01C60RLDQ

Brand : 1 Mad Fit

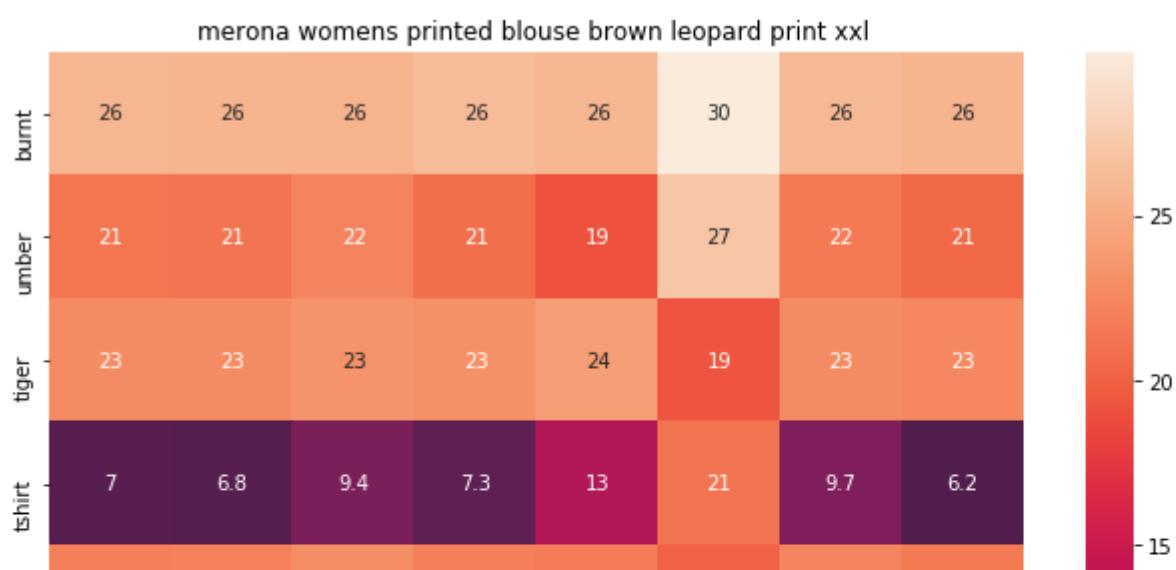
euclidean distance from input : 6.4634094

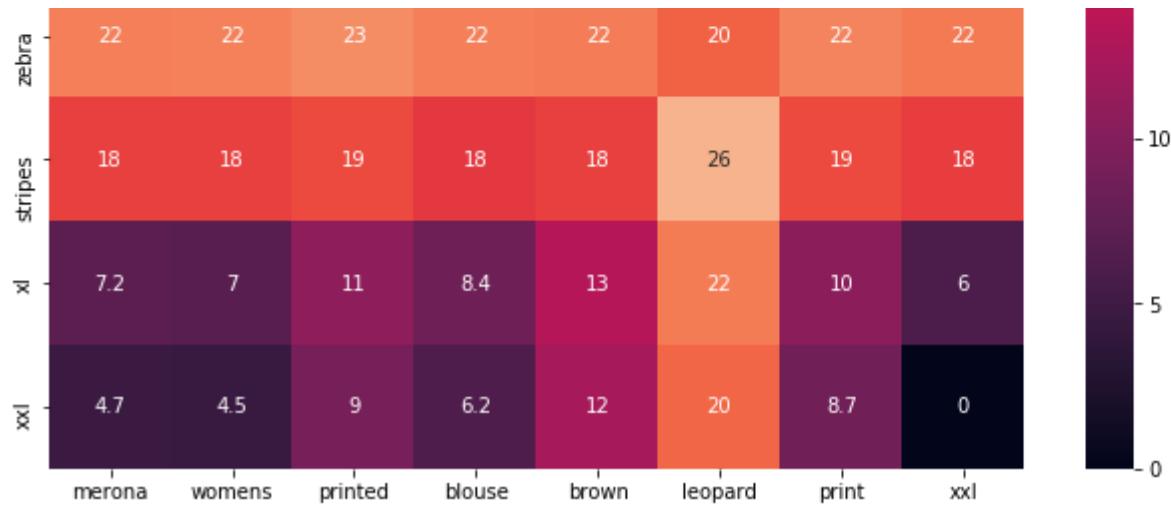


ASIN : B06XBY5QXL

Brand : Liz Claiborne

euclidean distance from input : 6.5392237





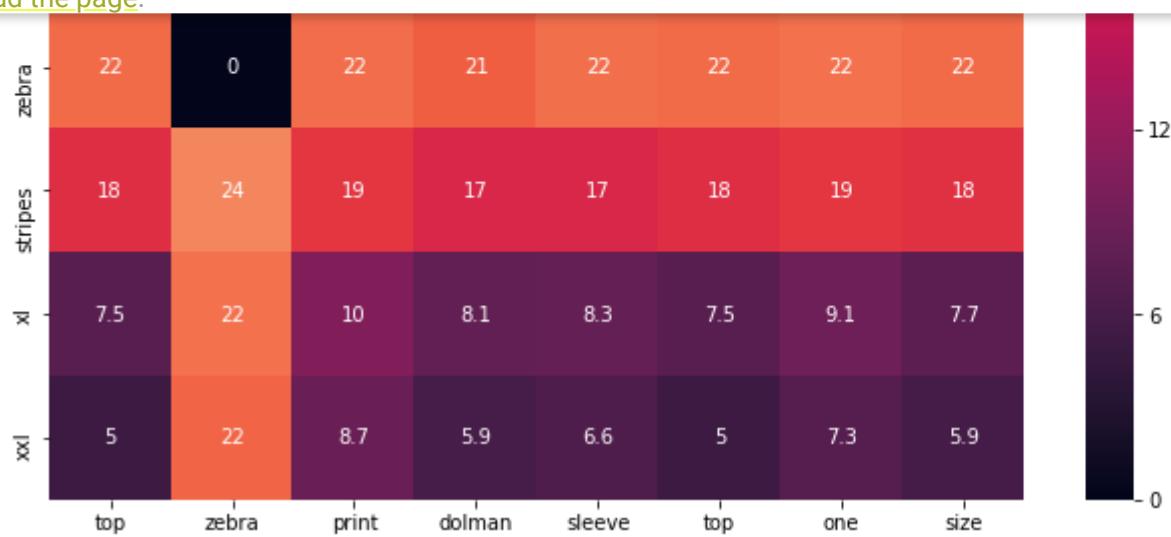
ASIN : B071YF3WDD

Brand : Merona

euclidean distance from input : 6.575504



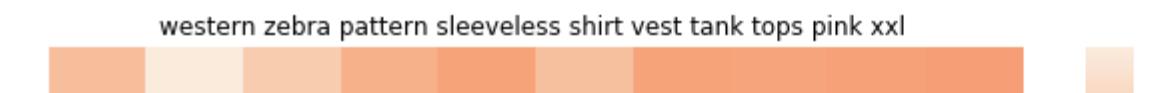
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

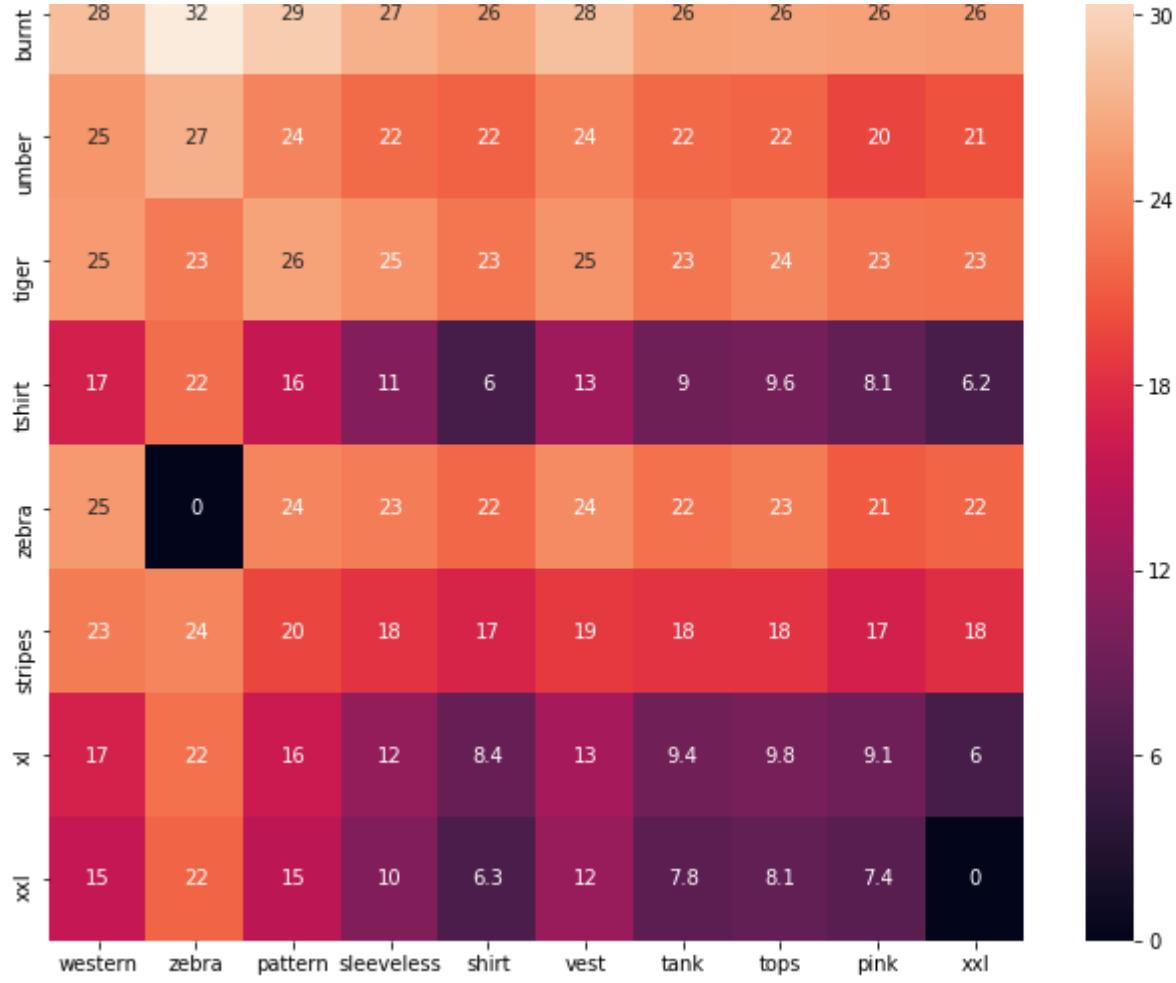


ASIN : B00H8A6ZLI

Brand : Vivian's Fashions

euclidean distance from input : 6.6382155





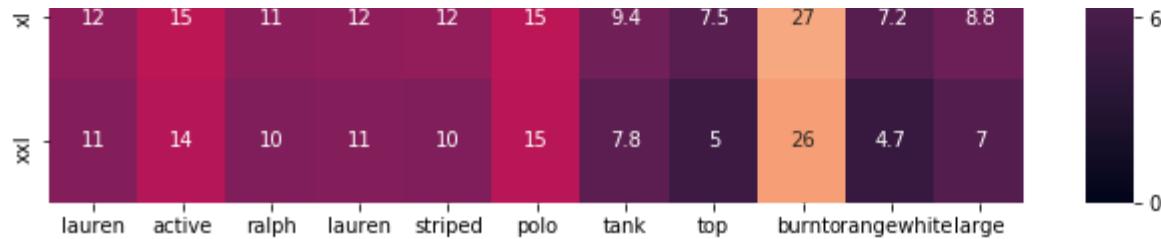
ASIN : B00Z6HEXWI

Brand : Black Temptation

euclidean distance from input : 6.660737

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

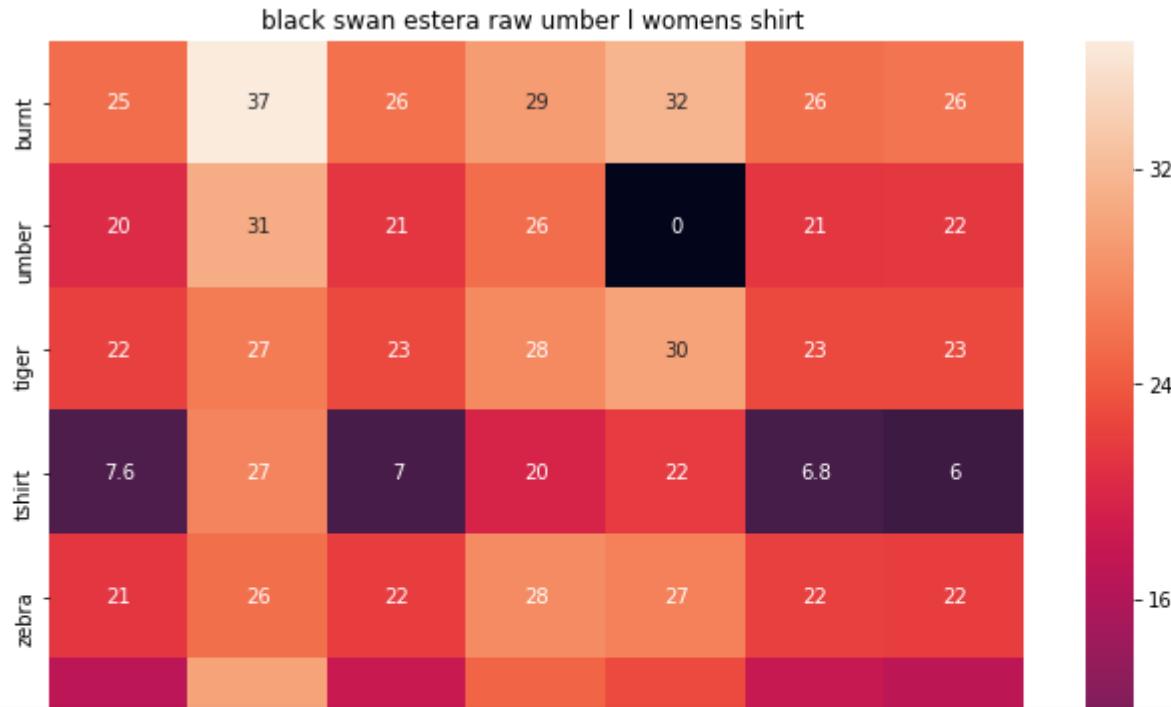




ASIN : B00ILGH50Y

Brand : Ralph Lauren Active

euclidean distance from input : 6.6839056



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

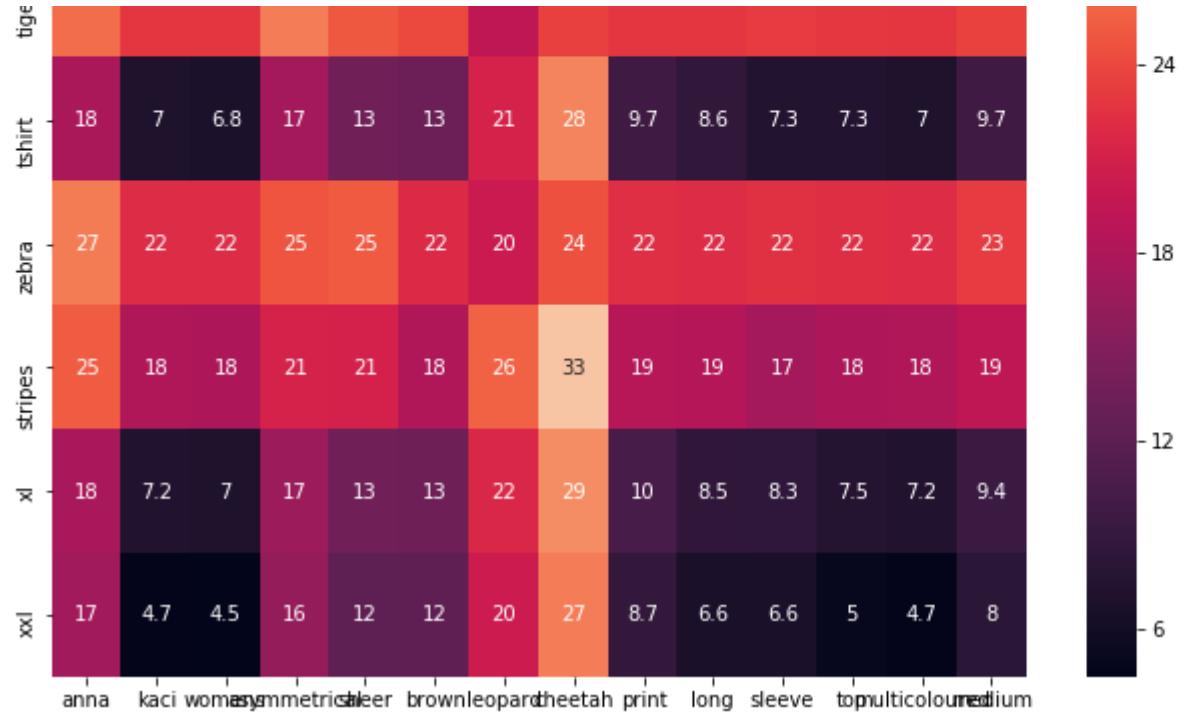


ASIN : B06Y1VN8WQ

Brand : Black Swan

euclidean distance from input : 6.7057643





ASIN : B00KSNTY7Y

Brand : Anna-Kaci

euclidean distance from input : 6.7061253

=====

▼ [9.6] Weighted similarity using brand and color.

some of the brand values are empty

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

```
# replace spaces with hyphen
brands = [x.replace(" ", "-") for x in data['brand'].values]
types = [x.replace(" ", "-") for x in data['product_type_name'].values]
colors = [x.replace(" ", "-") for x in data['color'].values]

brand_vectorizer = CountVectorizer()
brand_features = brand_vectorizer.fit_transform(brands)

type_vectorizer = CountVectorizer()
type_features = type_vectorizer.fit_transform(types)

color_vectorizer = CountVectorizer()
color_features = color_vectorizer.fit_transform(colors)

extra_features = hstack((brand_features, type_features, color_features)).tocsr()

def heat_map_w2v_brand(sentance1, sentance2, url, doc_id1, doc_id2, df_id1, df_id2, model):

    # sentance1 : title1, input apparel
```

```

# sentance2 : title2, recommended apparel
# url: apparel image url
# doc_id1: document id of input apparel
# doc_id2: document id of recommended apparel
# df_id1: index of document1 in the data frame
# df_id2: index of document2 in the data frame
# model: it can have two values, 1. avg 2. weighted

#s1_vec = np.array(#number_of_words_title1 * 300), each row is a vector(weighted/avg) of
s1_vec = get_word_vec(sentance1, doc_id1, model)
#s2_vec = np.array(#number_of_words_title2 * 300), each row is a vector(weighted/avg) of
s2_vec = get_word_vec(sentance2, doc_id2, model)

# s1_s2_dist = np.array(#number of words in title1 * #number of words in title2)
# s1_s2_dist[i,j] = euclidean distance between words i, j
s1_s2_dist = get_distance(s1_vec, s2_vec)

data_matrix = [['Asin', 'Brand', 'Color', 'Product type'],
               [data['asin'].loc[df_id1], brands[doc_id1], colors[doc_id1], types[doc_id1]], #
               [data['asin'].loc[df_id2], brands[doc_id2], colors[doc_id2], types[doc_id2]]] #

colorscale = [[0, '#1d004d'], [.5, '#f2e5ff'], [1, '#f2e5d1']] # to color the headings of e

# we create a table with the data_matrix
table = ff.create_table(data_matrix, index=True, colorscale=colorscale)
# plot it with plotly
plotly.offline.iplot(table, filename='simple_table')

# devide whole figure space into 25 * 1:10 grids

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```

# in first 25*10 grids we plot heatmap
ax1 = plt.subplot(gs[:, :-5])
# plotting the heatmap based on the pairwise distances
ax1 = sns.heatmap(np.round(s1_s2_dist, 6), annot=True)
# set the x axis labels as recommended apparels title
ax1.set_xticklabels(sentance2.split())
# set the y axis labels as input apparels title
ax1.set_yticklabels(sentance1.split())
# set title as recommended apparels title
ax1.set_title(sentance2)

# in last 25 * 10:15 grids we display image
ax2 = plt.subplot(gs[:, 10:16])
# we dont display grid lines and axis labels to images
ax2.grid(False)
ax2.set_xticks([])
ax2.set_yticks([])

# pass the url it display it

```

```
display_img(url, ax2, tig)
```

```
plt.show()
```

```
def idf_w2v_brand(doc_id, num_results, w1, w2, w3 = 0):
    # doc_id: apparel's id in given corpus
    # w1: weight for w2v features
    # w2: weight for brand and color features

    # pairwise_dist will store the distance from given input apparel to all remaining apparel
    # the metric we used here is cosine, the cosine distance is measured as  $K(X, Y) = \langle X, Y \rangle / \sqrt{\langle X, X \rangle \langle Y, Y \rangle}$ 
    # http://scikit-learn.org/stable/modules/metrics.html#cosine-similarity
    idf_w2v_dist = pairwise_distances(w2v_title_weight, w2v_title_weight[doc_id].reshape(1, -1))
    ex_feat_dist = pairwise_distances(extra_features, extra_features[doc_id])
    image_features_dist = pairwise_distances(bottleneck_features_train, bottleneck_features_test)

    pairwise_dist = (w1 * idf_w2v_dist + w2 * ex_feat_dist + w3 * image_features_dist) / float(w1 + w2 + w3)

    # np.argsort will return indices of 9 smallest distances
    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    # pdists will store the 9 smallest distances
    pdists = np.sort(pairwise_dist.flatten())[0:num_results]

    # data frame indices of the 9 smallest distance's
    df_indices = list(data.index[indices])

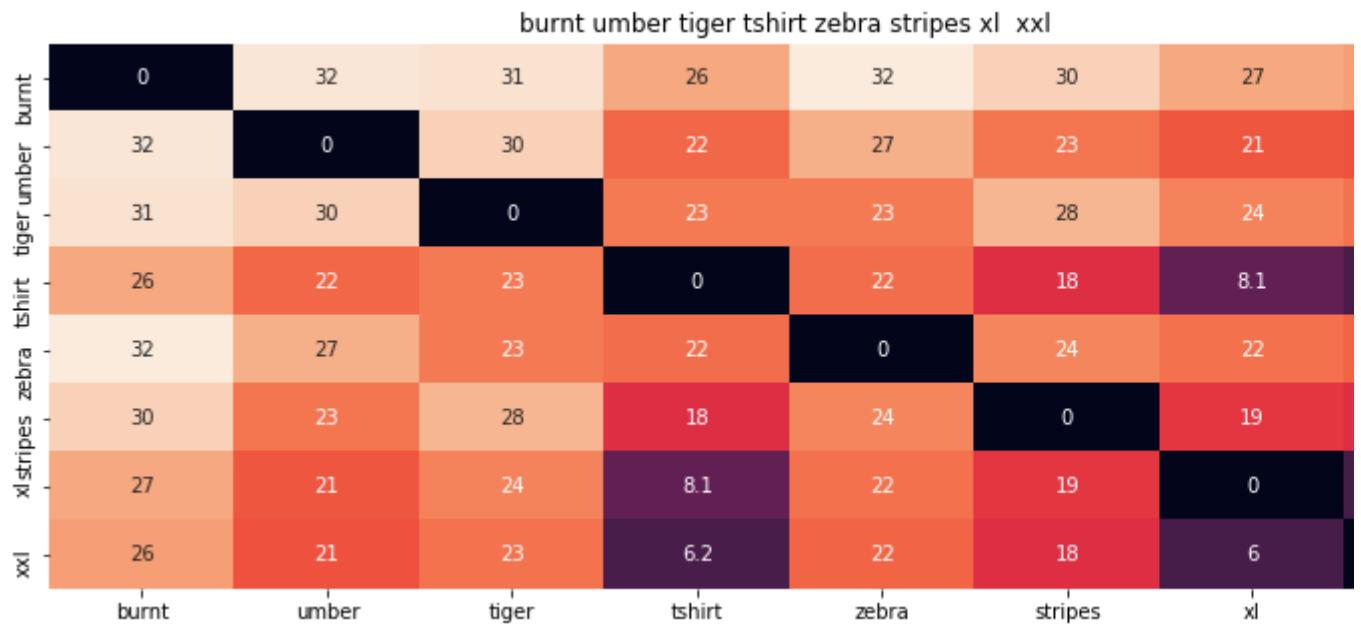
    for i in range(0, len(indices)):
        heat_map_w2v_brand(data['title'].loc[df_indices[0]], data['title'].loc[df_indices[i]],
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
        print('euclidean distance from input :', pdists[i])
        print('*'*125)
```

```
idf_w2v_brand(12566, 20, 5, 5)
# in the given heat map, each cell contains the euclidean distance between words i, j
```



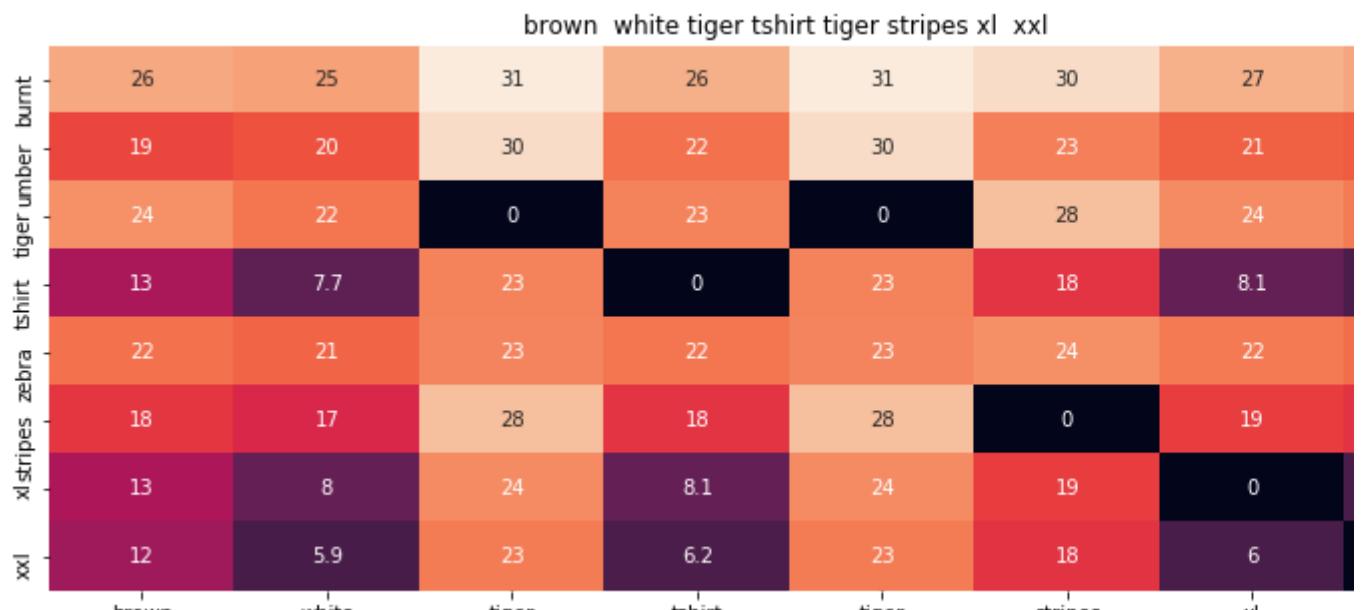


ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.001953125

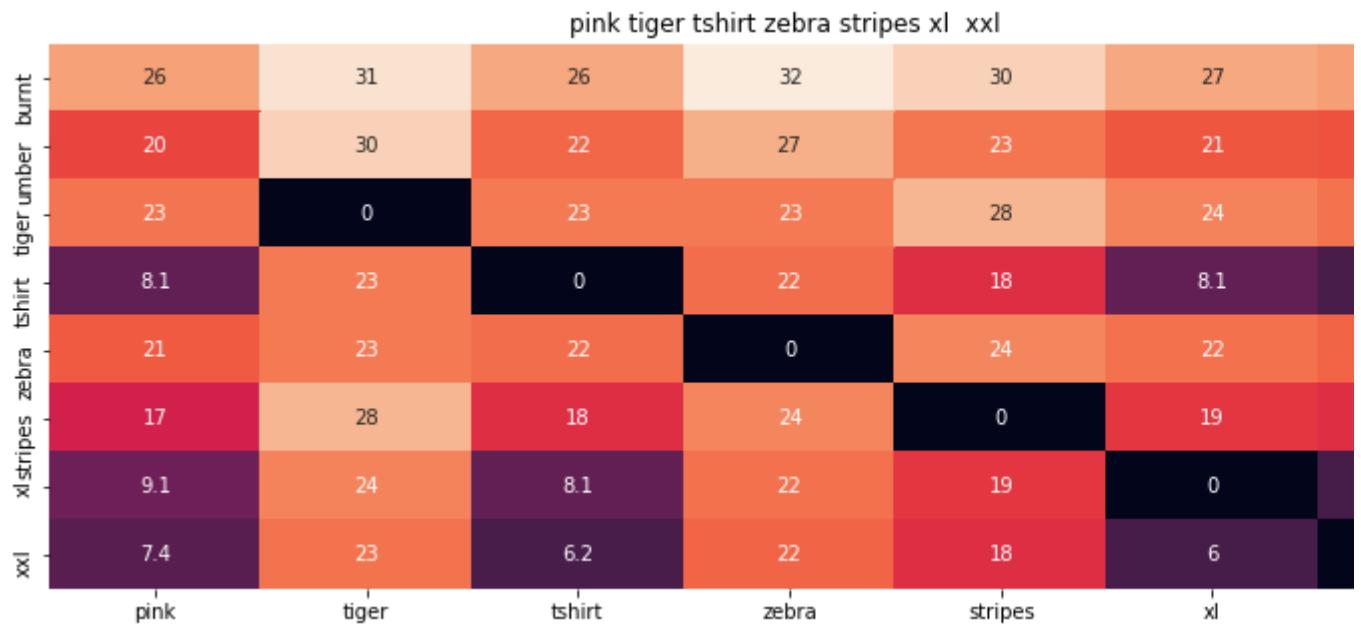
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B00JXQCWT0

Brand : Si Row

euclidean distance from input : 2.385471153259277



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



xxl	7.2	8	24	9.4	7.5	24	19	0
xxl	4.7	5.9	23	7.8	5	23	18	6

ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from input : 3.3871873857397703

	yellow	tiger	tank	top	tiger	stripes
burnt	27	31	26	26	31	1
tiger	20	30	22	21	30	1
umber	24	0	23	23	0	1
tshirt	13	23	9	7.3	23	1
zebra	21	23	22	22	23	1
stripes	17	28	18	18	28	1
xl	13	24	9.4	7.5	24	1

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQAUWA

Brand : Si Row

euclidean distance from input : 3.5518684389013915

	black	white	tiger	tank	top	tiger	stripes	1
burnt	25	25	31	26	26	31	1	
tiger	20	20	30	22	21	30	1	
umber	22	22	0	23	23	0	1	
tshirt	7.6	7.7	23	9	7.3	23	1	

zebra	21	21	23	22	22	23	
stripes	17	17	28	18	18	28	
xl	7.9	8	24	9.4	7.5	24	
xxl	5.9	5.9	23	7.8	5	23	
	black	white	tiger	tank	top	tiger	

ASIN : B00JXQA094

Brand : Si Row

euclidean distance from input : 3.5536182405371335

	yellow	tiger	tshirt	tiger	stripes	1
burnt	27	31	26	31		
umber	20	30	22	30		
tiger	24	0	23	0		
tshirt	13	23	0	23		
xxl	21	23	22	23		
	yellow	tiger	tshirt	tiger	stripes	1

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

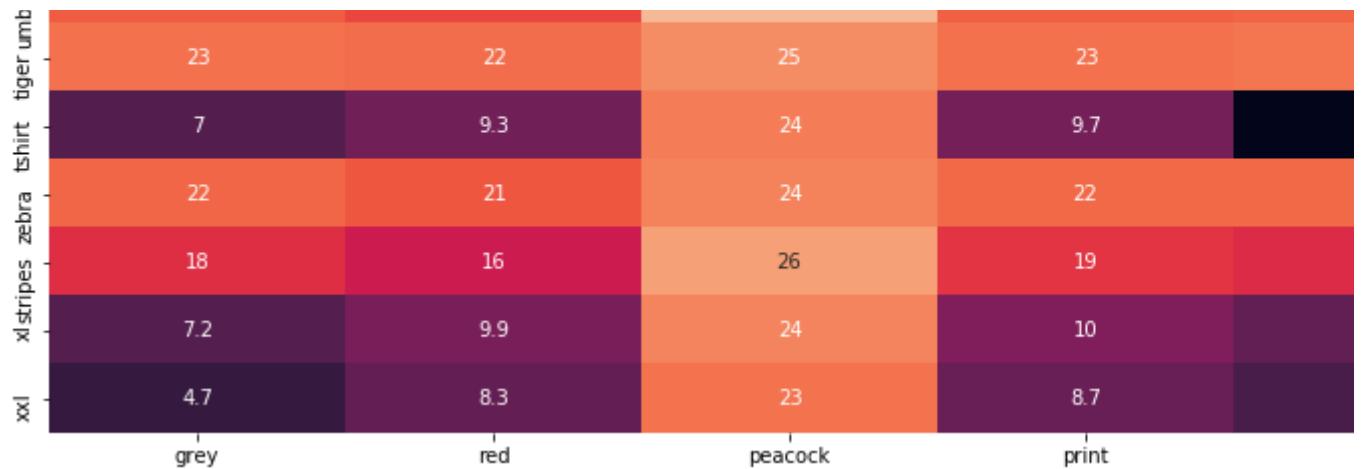
yellow	13	24	8.1	24	
tiger	12	23	6.2	23	
xxl					

ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from input : 3.653828048886743

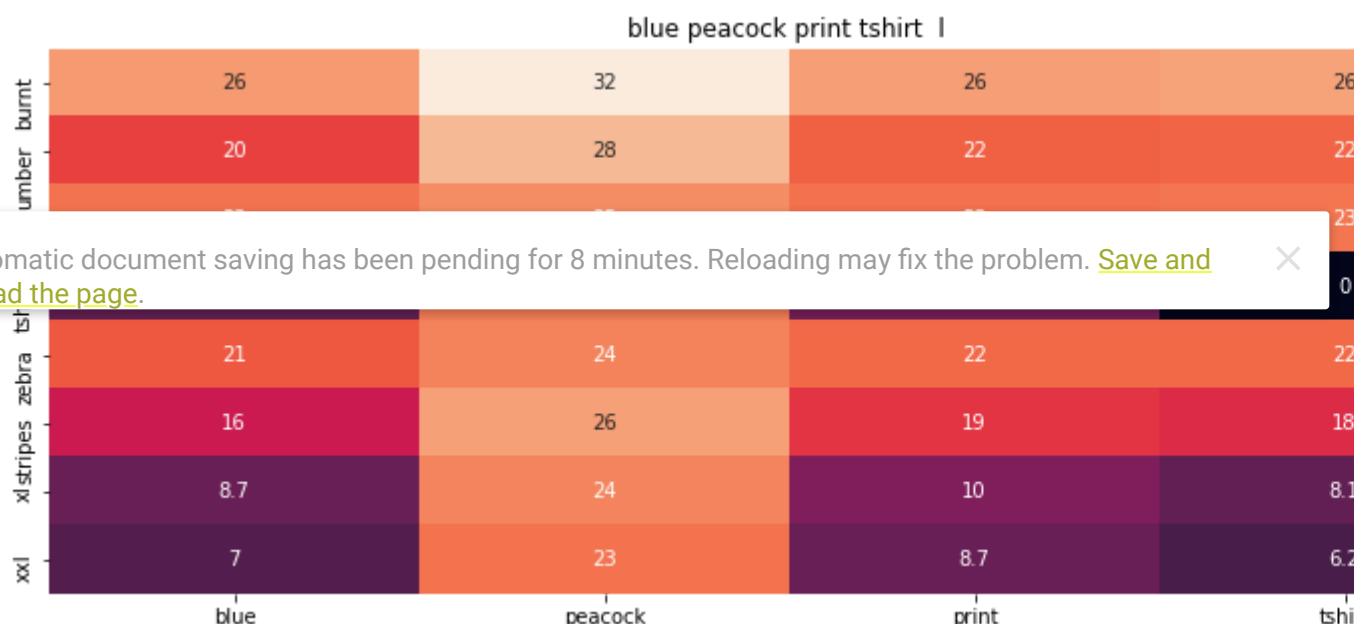
	grey	red	peacock	print	tshirt	1
burnt	26	25	32	26		
xxl	21	20	28	22		



ASIN : B00JXQCFRS

Brand : Si Row

euclidean distance from input : 4.128811645688501



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X 0

ASIN : B00JXQC8L6

Brand : Si Row

euclidean distance from input : 4.203900146665063

	red	butterfly	black	white	tank	top	xl
burnt	25	33	25	25	26	26	27
tiger umber	20	27	20	20	22	21	21
tiger	22	27	22	22	23	23	24
tshirt	9.3	21	7.6	7.7	9	7.3	8.1
zebra	21	26	21	21	22	22	22
stripes	16	26	17	17	18	18	19
xxl	9.9	22	7.9	8	9.4	7.5	0
xl	8.3	21	5.9	5.9	7.8	5	6

ASIN : B00JV63CW2

Brand : Si Row

euclidean distance from input : 4.286586761655298

=====

kingde black white zebra stripe sleeveless vestbqn38

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	kingde	black	white	zebra	stripe	sleeveless
tiger umber	23	22	22	23	26	25
tshirt	7	7.6	7.7	22	14	11
zebra	22	21	21	0	24	23
stripes	18	17	17	24	15	18
xxl	7.2	7.9	8	22	15	12
xl	4.7	5.9	5.9	22	14	10

ASIN : B015H41F6G

Brand : KINGDE

euclidean distance from input : 4.389370597243721

=====

	yellow	pink	butterfly	color	burst	tshirt	xl
burnt	27	26	33	27	32	26	27
tiger	20	20	27	20	32	22	21
umber	24	23	27	24	33	23	24
tshirt	13	8.1	21	13	26	0	8.1
zebra	21	21	26	22	32	22	22
stripes	17	17	26	17	29	18	19
xl	13	9.1	22	13	26	8.1	0
xxl	12	7.4	21	12	25	6.2	6
	yellow	pink	butterfly	color	burst	tshirt	xl

ASIN : B00JXQBBMI

Brand : Si Row

euclidean distance from input : 4.397909927548852

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	lsu	tigers	colosseum	womens	yellow	purple	slit	back	12	sleeves	tshirt
burnt	30	36	31	26	27	26	29	27	26	29	
tiger	24	34	28	21	20	19	29	23	21	24	
umber	26	17	28	23	24	24	30	23	23	27	
tshirt	16	29	21	6.8	13	10	23	10	7	14	
zebra	24	27	27	22	21	21	30	23	22	26	
stripes	23	32	26	18	17	16	27	19	18	19	
xl	16	30	21	7	13	11	24	11	7.2	15	
xxl	15	29	20	4.5	12	10	23	8.8	4.7	14	
	lsu	tigers	colosseum	womens	yellow	purple	slit	back	12	sleeves	tshirt

ASIN : B073R5Q8HD

Brand : Colosseum

euclidean distance from input : 4.451228583693917

=====

	edista	embellished	zebra	womens	small	tank	cami
burnt	26	28	32	26	26	26	28
tiger	21	22	27	21	21	22	22
umber	23	25	23	23	23	23	25
tshirt	7	15	22	6.8	8.5	9	11
zebra	22	24	0	22	22	22	23
stripes	18	20	24	18	19	18	19
xl	7.2	15	22	7	8.4	9.4	12
xxl	4.7	14	22	4.5	6.4	7.8	11

ASIN : B074P8MD22

Brand : Edista

euclidean distance from input : 4.518977797866279

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	red	pink	floral	heel	sleeveless	shirt	xl
burnt	25	26	28	35	27	26	27
tiger	20	20	21	31	22	22	21
umber	22	23	24	32	25	23	24
tshirt	9.3	8.1	12	24	11	6	8.1
zebra	21	21	23	29	23	22	22
stripes	16	17	19	27	18	17	19
xl	9.9	9.1	12	25	12	8.4	0
xxl	8.3	7.4	11	24	10	6.3	6

ASIN : B00JV63Q0E

Brand : Si Row

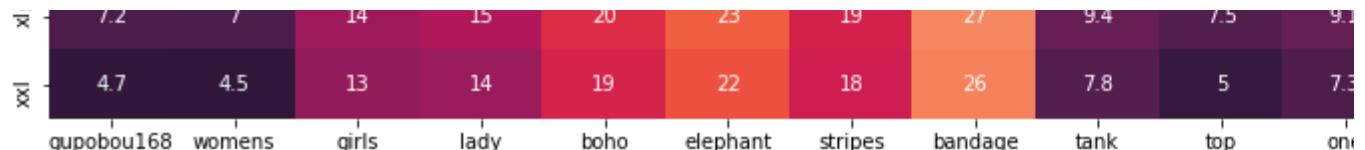
euclidean distance from input : 4.52937545794436

	stanzino womens zebra print dolman sleeve chiffon top teal								
burnt	26	26	32	26	26	26	29	26	26
tiger	21	21	27	22	20	22	21	21	21
umber	23	23	23	23	23	23	27	23	23
tshirt	7	6.8	22	9.7	7.8	7.3	15	7.3	7.3
zebra	22	22	0	22	21	22	24	22	22
stripes	18	18	24	19	17	17	20	18	18
xxl	7.2	7	22	10	8.1	8.3	16	7.5	7.5
	4.7	4.5	22	8.7	5.9	6.6	15	5	5
	stanzino	womens	zebra	print	dolman	sleeve	chiffon	top	

ASIN : B00C0I3U3E

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

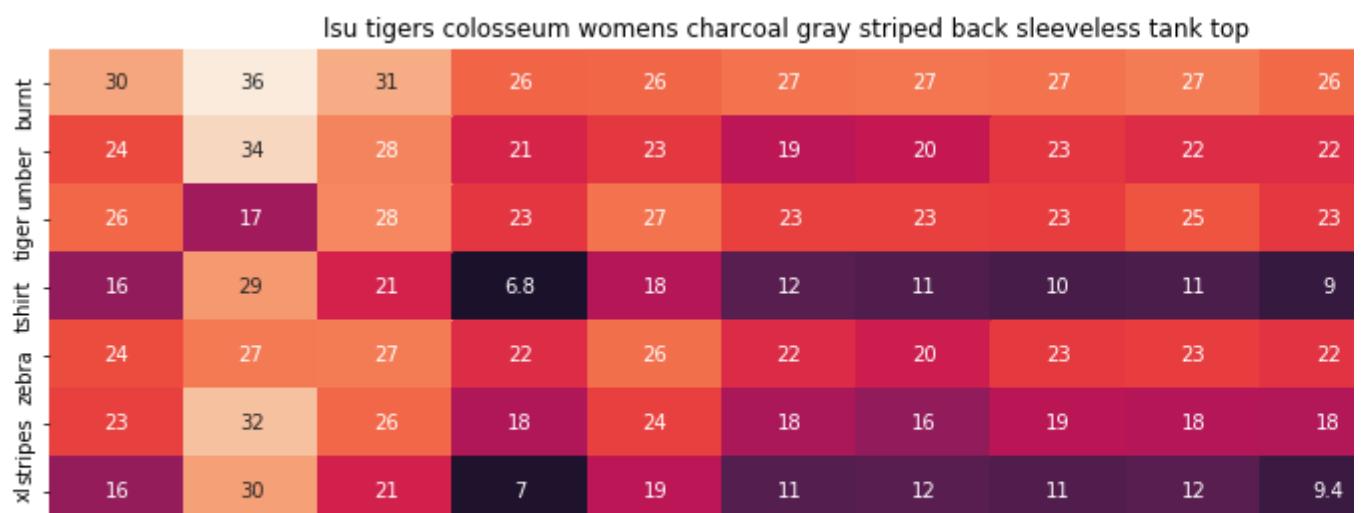
	gupobou168 womens girls lady boho elephant stripes bandage tank top one size											
burnt	26	26	28	28	31	32	30	35	26	26	26	26
tiger	21	21	24	24	25	29	23	32	22	21	22	22
umber	23	23	25	25	29	18	28	33	23	23	23	23
tshirt	7	6.8	14	15	19	23	18	25	9	7.3	8.8	8.8
zebra	22	22	25	24	27	22	24	32	22	22	22	22
stripes	18	18	22	23	24	28	0	28	18	18	18	19
xxl	7.2	7	24	25	26	27	28	27	7.5	7.5	7.5	7.5
	stanzino	womens	zebra	print	dolman	sleeve	chiffon	top				



ASIN : B01ER18406

Brand : GuPoBoU168

euclidean distance from input : 4.546817024028215

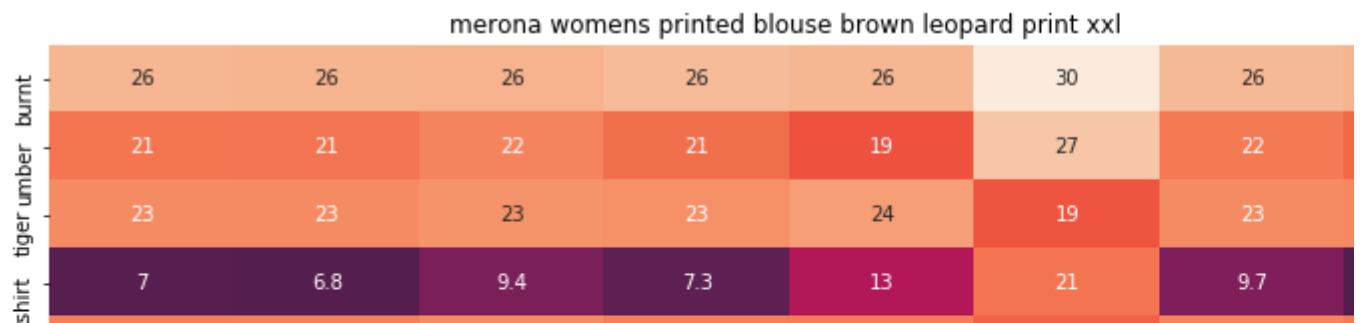


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).

ASIN : B073R4ZM7Y

Brand : Colosseum

euclidean distance from input : 4.548355925918037

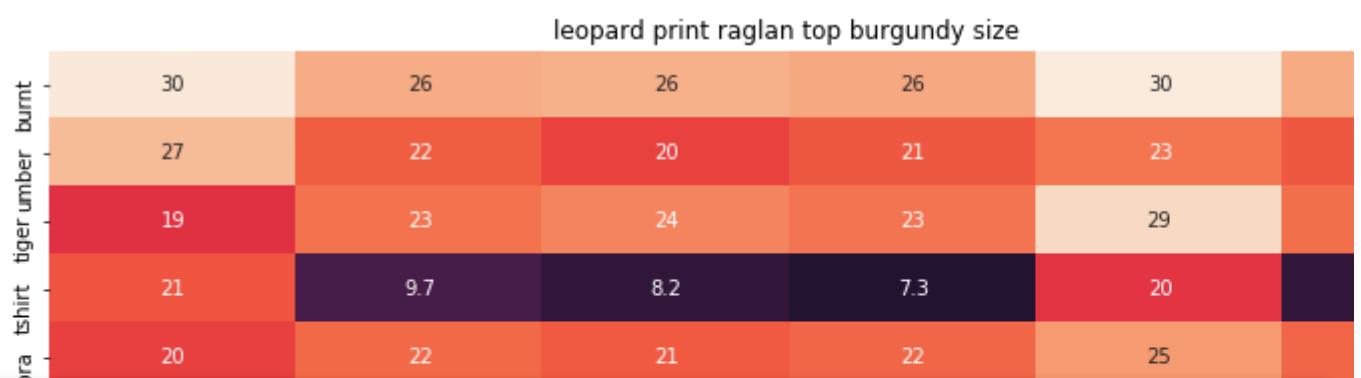




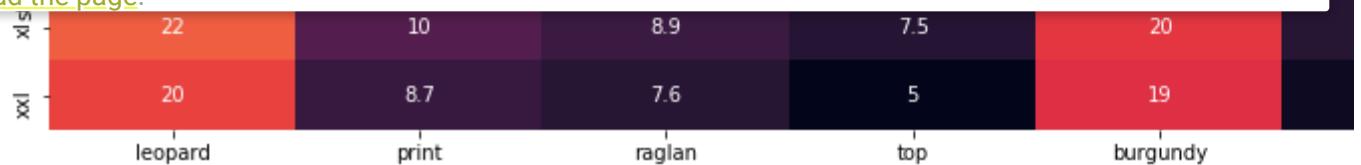
ASIN : B071YF3WDD

Brand : Merona

euclidean distance from input : 4.610627425551827



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B01C60RLDQ

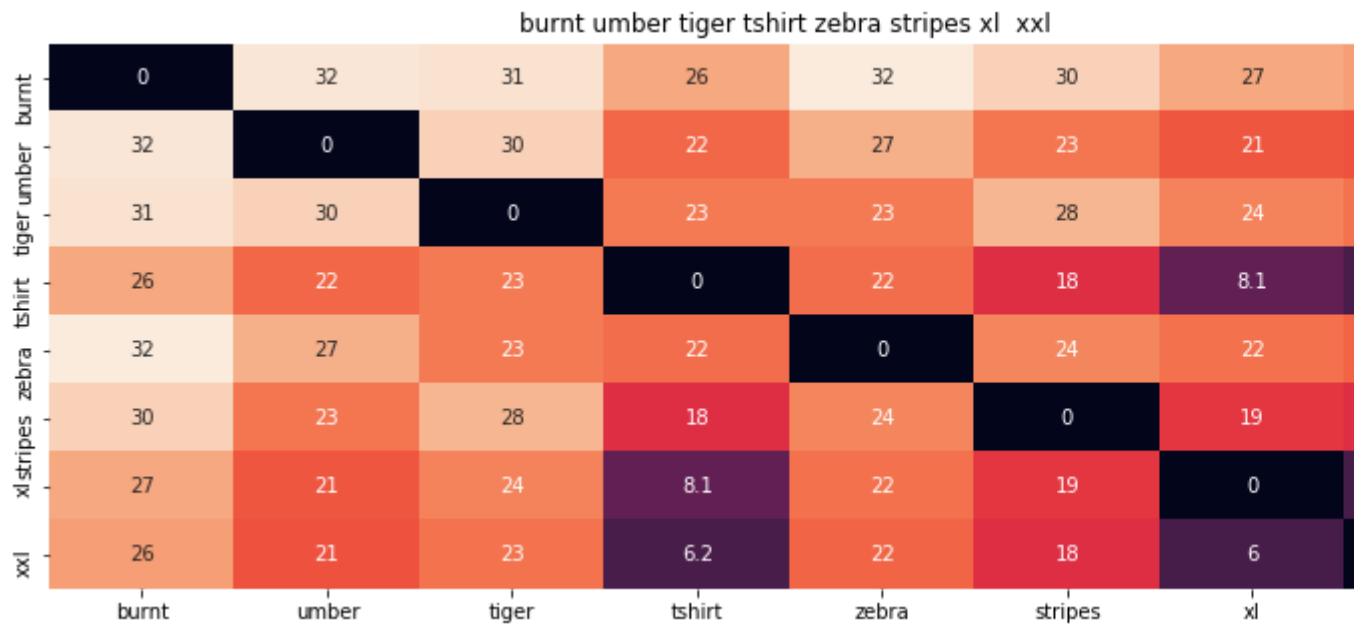
Brand : 1 Mad Fit

euclidean distance from input : 4.645918274287157

```
# brand and color weight =50
# title vector weight = 5
```

```
idf_w2v_brand(12566, 5, 50, 20)
```



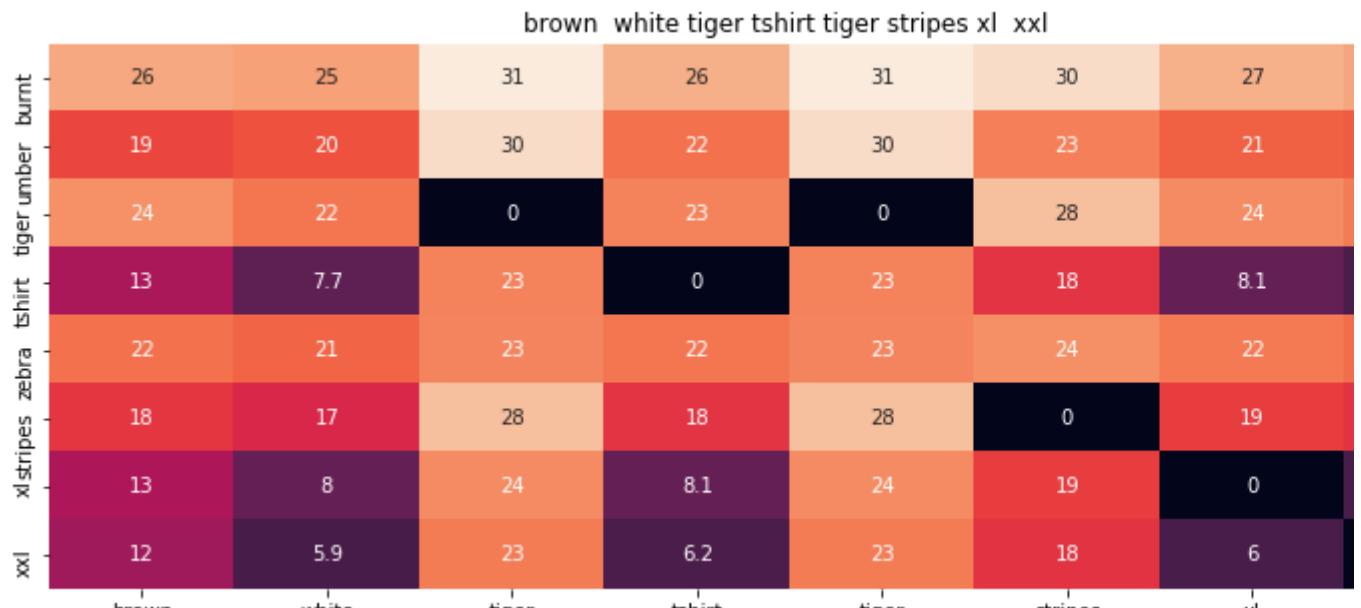


ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.0003551136363636364

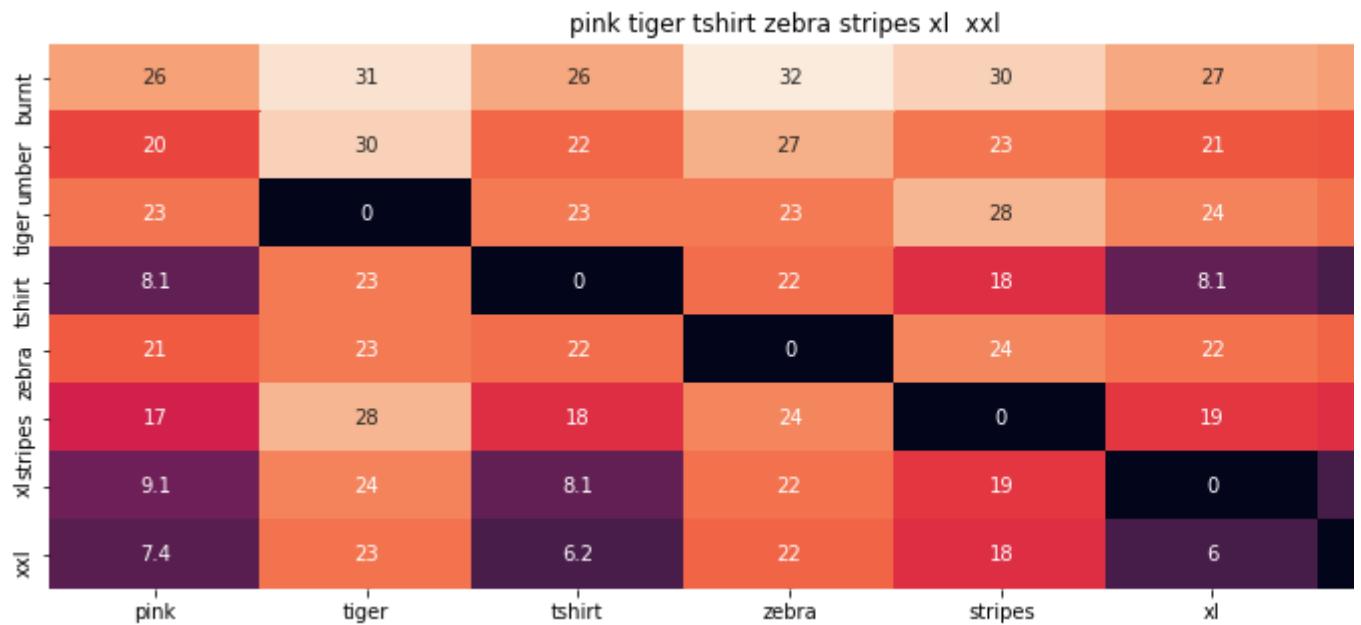
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B00JXQCWT0

Brand : Si Row

euclidean distance from input : 0.43372202786532316



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



xxl	7.2	8	24	9.4	7.5	24	19	0
xxl	4.7	5.9	23	7.8	5	23	18	6

ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from input : 1.7729360757124906

	yellow	tiger	tank	top	tiger	stripes
burnt	27	31	26	26	31	1
tiger	20	30	22	21	30	1
umber	24	0	23	23	0	1
tshirt	13	23	9	7.3	23	1
zebra	21	23	22	22	23	1
stripes	17	28	18	18	28	1
xl	13	24	9.4	7.5	24	1

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQAUWA

Brand : Si Row

euclidean distance from input : 1.8028780853782398

	black	white	tiger	tank	top	tiger	stripes	1
burnt	25	25	31	26	26	31	1	
tiger	20	20	30	22	21	30	1	
umber	22	22	0	23	23	0	1	
tshirt	7.6	7.7	23	9	7.3	23	1	

zebra	21	21	23	22	22	23	
stripes	17	17	28	18	18	28	
xl	7.9	8	24	9.4	7.5	24	
xxl	5.9	5.9	23	7.8	5	23	
	black	white	tiger	tank	top	tiger	

ASIN : B00JXQA094

Brand : Si Row

euclidean distance from input : 1.803196231130193

	yellow	tiger	tshirt	tiger	stripes	1
burnt	27	31	26	31		
umber	20	30	22	30		
tiger	24	0	23	0		
tshirt	13	23	0	23		
xxl	21	23	22	23		
	yellow	tiger	tshirt	tiger	stripes	1

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

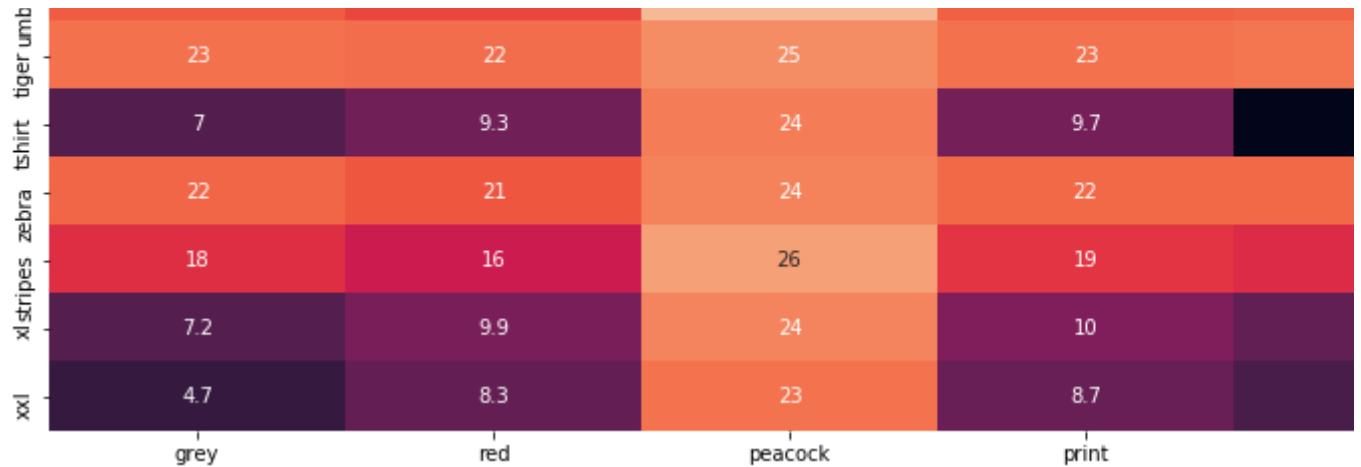
stripes	13	24	8.1	24	
xxl	12	23	6.2	23	
	yellow	tiger	tshirt	tiger	
	yellow	tiger	tshirt	tiger	
	yellow	tiger	tshirt	tiger	

ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from input : 1.8214161962846673

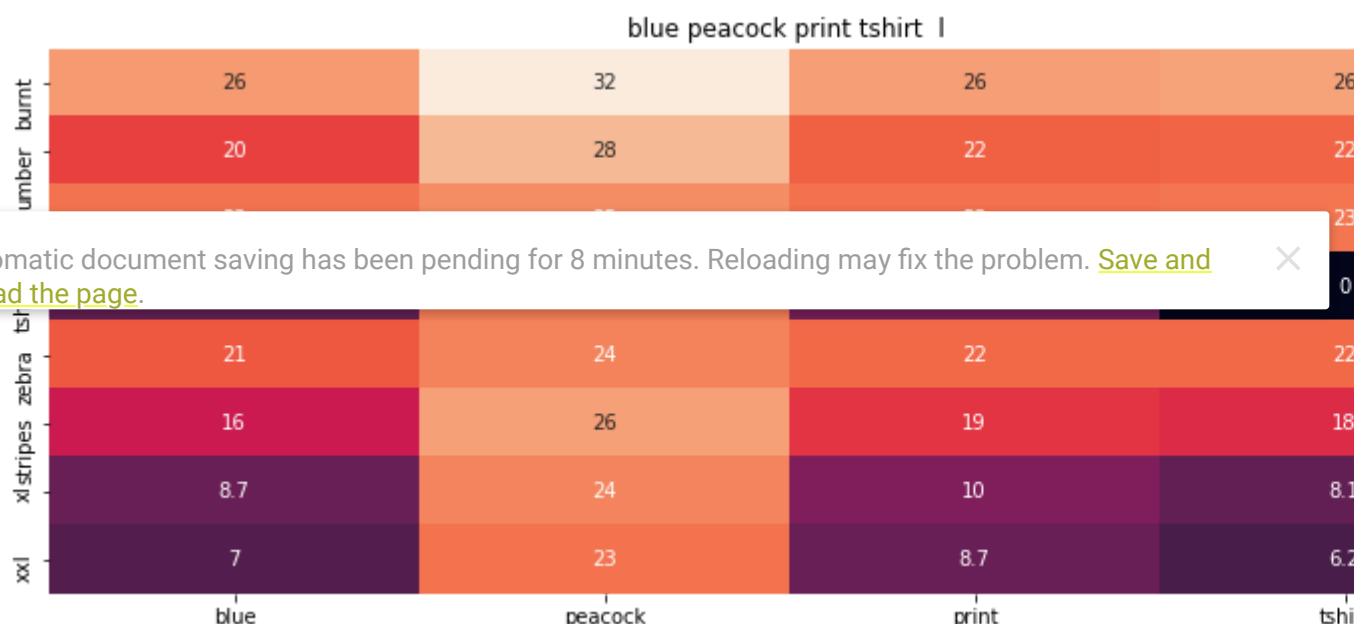
	grey	red	peacock	print	tshirt	1
burnt	26	25	32	26		
xxl	21	20	28	22		
	grey	red	peacock	print	tshirt	1
	grey	red	peacock	print	tshirt	1



ASIN : B00JXQCFRS

Brand : Si Row

euclidean distance from input : 1.9077768502486234



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X 0

ASIN : B00JXQC8L6

Brand : Si Row

euclidean distance from input : 1.9214293049716347

	red	butterfly	black	white	tank	top	xl
burnt	25	33	25	25	26	26	27
tiger	20	27	20	20	22	21	21
umbre	22	27	22	22	23	23	24
tshirt	9.3	21	7.6	7.7	9	7.3	8.1
zebra	21	26	21	21	22	22	22
stripes	16	26	17	17	18	18	19
xxl	9.9	22	7.9	8	9.4	7.5	0
xl	8.3	21	5.9	5.9	7.8	5	6

ASIN : B00JV63CW2

Brand : Si Row

euclidean distance from input : 1.9364632349698592

=====

yellow pink butterfly color burst tshirt xl xxl

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	yellow	pink	butterfly	color	burst	tshirt	xl
tiger	24	23	27	24	33	23	24
umbre	13	8.1	21	13	26	0	8.1
zebra	21	21	26	22	32	22	22
stripes	17	17	26	17	29	18	19
xxl	13	9.1	22	13	26	8.1	0
xl	12	7.4	21	12	25	6.2	6

ASIN : B00JXQBBMI

Brand : Si Row

euclidean distance from input : 1.956703810586869

=====

	red	pink	floral	heel	sleeveless	shirt	xl
burnt	25	26	28	35	27	26	27
tiger	20	20	21	31	22	22	21
umber	22	23	24	32	25	23	24
tshirt	9.3	8.1	12	24	11	6	8.1
zebra	21	21	23	29	23	22	22
stripes	16	17	19	27	18	17	19
xl	9.9	9.1	12	25	12	8.4	0
xxl	8.3	7.4	11	24	10	6.3	6

ASIN : B00JV63Q0E

Brand : Si Row

euclidean distance from input : 1.9806066342951432

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	purple	floral	heel	sleeveless	shirt	xl
burnt	26	28	35	27	26	27
tiger	19	21	31	22	22	21
umber	24	24	32	25	23	24
tshirt	10	12	24	11	6	8.1
zebra	21	23	29	23	22	22
stripes	16	19	27	18	17	19
xl	11	12	25	12	8.4	0
xxl	10	11	24	10	6.3	6

ASIN : B00JV63VC8

Brand : Si Row

euclidean distance from input : 2.0121855999157043

=====

	pink	rose	bud	heel	dalmatian	tank	top	xl
burnt	26	28	39	35	35	26	26	27
tiger	20	25	35	31	31	22	21	21
umber	23	27	36	32	27	23	23	24
tshirt	8.1	18	31	24	26	9	7.3	8.1
zebra	21	26	36	29	26	22	22	22
stripes	17	23	35	27	29	18	18	19
xxl	9.1	17	31	25	27	9.4	7.5	0
	7.4	16	31	24	26	7.8	5	6
pink								
rose								
bud								
heel								
dalmatian								
tank								
top								
xl								

ASIN : B00JXQAX2C

Brand : Si Row

euclidean distance from input : 2.013351787548872

	blue	green	butterfly	color	burst	t	shirt	l
burnt	26	26	33	27	32			
tiger	20	21	27	20	32			
umber	23	23	27	24	33			
tshirt	8	11	21	13	26			
zebra	21	23	26	22	32			
stripes	16	17	26	17	29			
xxl	8.7	12	22	13	26			
	7	10	21	12	25			
blue								
green								
butterfly								
color								
burst								
t								
shirt								
l								

ASIN : B00JXQC0C8

Brand : Si Row

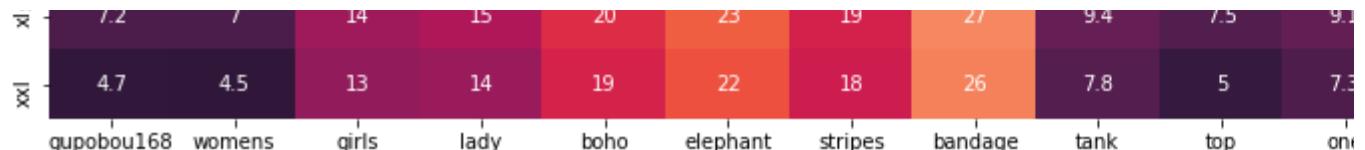
euclidean distance from input : 2.013883348273304

	red	rose	bud	heel	dalmatian	tshirt	xl	xxl
burnt	25	28	39	35	35	26	27	
tiger	20	25	35	31	31	22	21	
umber	22	27	36	32	27	23	24	
tshirt	9.3	18	31	24	26	0	8.1	
zebra	21	26	36	29	26	22	22	
stripes	16	23	35	27	29	18	19	
xl	9.9	17	31	25	27	8.1	0	
xxl	8.3	16	31	24	26	6.2	6	
	red	rose	bud	heel	dalmatian	tshirt	xl	xxl

ASIN : B00JXQABB0

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

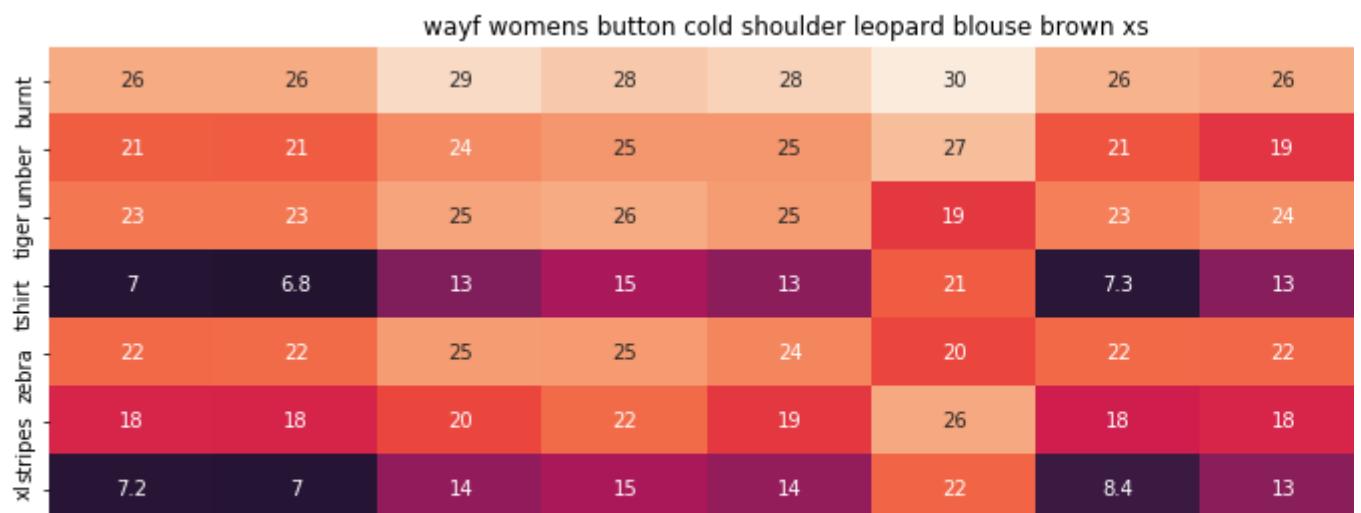
	gupobou	168	womens	girls	lady	boho	elephant	stripes	bandage	tank	top	one size
burnt	26	26	28	28	31	32	30	35	26	26	26	
tiger	21	21	24	24	25	29	23	32	22	21	22	
umber	23	23	25	25	29	18	28	33	23	23	23	
tshirt	7	6.8	14	15	19	23	18	25	9	7.3	8.8	
zebra	22	22	25	24	27	22	24	32	22	22	22	
stripes	18	18	22	23	24	28	0	28	18	18	19	
xxl	7.2	7	11	15	20	27	18	27	8.4	7.5	8.1	
	gupobou	168	womens	girls	lady	boho	elephant	stripes	bandage	tank	top	one size



ASIN : B01ER18406

Brand : GuPoBoU168

euclidean distance from input : 2.6562041677776853

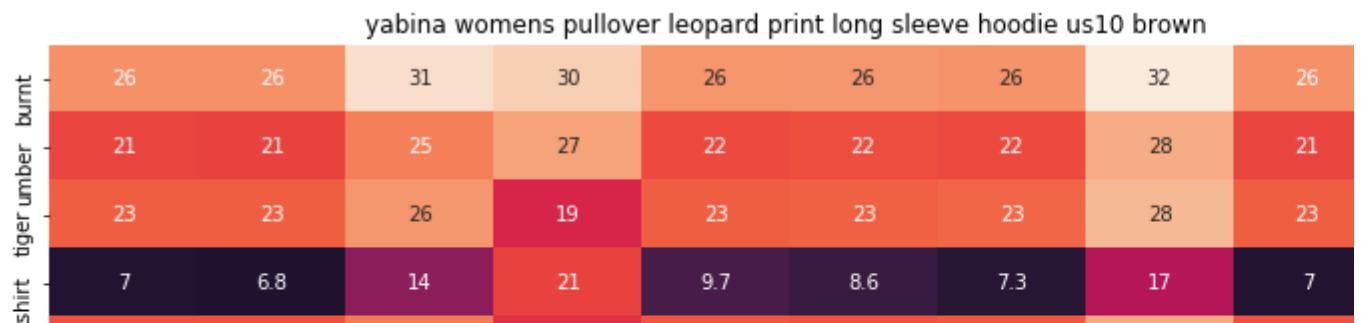


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).

ASIN : B01LZ7BQ4H

Brand : WAYF

euclidean distance from input : 2.684906782301833



zebra	22	22	25	20	22	22	22	27	22
stripes	18	18	21	26	19	19	17	23	18
xl	7.2	7	16	22	10	8.5	8.3	19	7.2
xxl	4.7	4.5	15	20	8.7	6.6	6.6	18	4.7

ASIN : B01KJUM6JI

Brand : YABINA

euclidean distance from input : 2.6858381926578345

	wayf	womens	small	button	leopard	cold	shoulder	blouse	brown
burnt	26	26	26	29	30	28	28	26	
tiger	21	21	21	24	27	25	25	21	
umber	23	23	23	25	19	26	25	23	
tshirt	7	6.8	8.5	13	21	15	13	7.3	
way	22	22	22	25	20	25	24	22	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

wayf	7.2	7	8.4	14	22	15	14	8.4
xxl	4.7	4.5	6.4	12	20	14	12	6.2

ASIN : B01M06V4X1

Brand : WAYF

euclidean distance from input : 2.694761948650377

▼ [10.2] Keras and Tensorflow to extract features

```
import numpy as np
from keras.preprocessing.image import ImageDataGenerator
from keras.models import Sequential
from keras.layers import Dropout, Flatten, Dense
from keras import applications
```

```
from __future__ import print_function
from sklearn.metrics import pairwise_distances
import matplotlib.pyplot as plt
import requests
from PIL import Image
import pandas as pd
import pickle
```

Using TensorFlow backend.

```
# https://gist.github.com/fchollet/f35fbc80e066a49d65f1688a7e99f069
# Code reference: https://blog.keras.io/building-powerful-image-classification-models-using-vgg16-and-vgg19.html
```

```
# This code takes 40 minutes to run on a modern GPU (graphics card)
# like Nvidia 1050.
# GPU (Nvidia 1050): 0.175 seconds per image
```

```
# This code takes 160 minutes to run on a high end i7 CPU
# CPU (i7): 0.615 seconds per image.
```

```
#Do NOT run this code unless you want to wait a few hours for it to generate output
```

```
# each image is converted into 25088 length dense-vector
```

```
...
```

```
# dimensions of our images.
```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
train_data_dir = 'images2/'
nb_train_samples = 16042
epochs = 50
batch_size = 1
```

```
def save_bottlebeck_features():
```

```
#Function to compute VGG-16 CNN for image feature extraction.
```

```
asins = []
datagen = ImageDataGenerator(rescale=1. / 255)

# build the VGG16 network
model = applications.VGG16(include_top=False, weights='imagenet')
generator = datagen.flow_from_directory(
    train_data_dir,
    target_size=(img_width, img_height),
    batch_size=batch_size,
    class_mode=None,
```

```

shuffle=False)

for i in generator.filenames:
    asins.append(i[2:-5])

bottleneck_features_train = model.predict_generator(generator, nb_train_samples // batch_
bottleneck_features_train = bottleneck_features_train.reshape((16042,25088))

np.save(open('16k_data_cnn_features.npy', 'wb'), bottleneck_features_train)
np.save(open('16k_data_cnn_feature_asins.npy', 'wb'), np.array(asins))

save_bottlebeck_features()

...

```

👤 "\n# dimensions of our images.\nimg_width, img_height = 224, 224\n\nmodel_weights_pa

▼ [10.3] Visual features based product similarity.

```

#load the features and corresponding ASINS info.
bottleneck_features_train = np.load('16k_data_cnn_features.npy')
asins = np.load('16k_data_cnn_feature_asins.npy')
asins = list(asins)

```

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

```
from IPython.display import display, Image, SVG, Math, YouTubeVideo
```

```

#get similar products using CNN features (VGG-16)
def get_similar_products_cnn(doc_id, num_results):
    doc_id = asins.index(df_asins[doc_id])
    pairwise_dist = pairwise_distances(bottleneck_features_train, bottleneck_features_train[d

    indices = np.argsort(pairwise_dist.flatten())[0:num_results]
    pdists = np.sort(pairwise_dist.flatten())[0:num_results]

    for i in range(len(indices)):
        rows = data[['medium_image_url', 'title']].loc[data['asin'] == asins[indices[i]]]
        for idx, row in rows.iterrows():
            display(Image(url=row['medium_image_url'], embed=True))
            print('Product Title: ', row['title'])
            print('Euclidean Distance from input image: ', pdists[i])
            print('Amazon Url: www.amazon.com/dp/' + asins[indices[i]])

```

```
get_similar_products_cnn(12566, 20)
```



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



Product Title: burnt umber tiger tshirt zebra stripes xl xxl

Euclidean Distance from input image: 0.044194173

Amazon Url: www.amazon.com/dp/B00JXQB5FQ



Product Title: pink tiger tshirt zebra stripes xl xxl

Euclidean Distance from input image: 30.050056

Amazon Url: www.amazon.com/dp/B00JXQASS6



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and](#)

[reload the page](#).

Product Title: yellow tiger tshirt tiger stripes xl

Euclidean Distance from input image: 41.261112

Amazon Url: www.amazon.com/dp/B00JXQCUIC



Product Title: brown white tiger tshirt tiger stripes xl xxl

Euclidean Distance from input image: 44.0002

Amazon Url: www.amazon.com/dp/B00JXQCWT0



Product Title: kawaii pastel tops tees pink flower design
Euclidean Distance from input image: 47.38251
Amazon Url: www.amazon.com/dp/B071FCWD97



Product Title: womens thin style tops tees pastel watermelon print
Euclidean Distance from input image: 47.71839
Amazon Url: www.amazon.com/dp/B01JUNHBRM



Product Title: kawaii pastel tops tees baby blue flower design
Euclidean Distance from input image: 47.9021
Amazon Url: www.amazon.com/dp/B071SBCY9W



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

Product Title: edv cheetah run purple multi xl
Euclidean Distance from input image: 48.046467
Amazon Url: www.amazon.com/dp/B01CUPYBM0



Product Title: danskin womens vneck loose performance tee xsmall pink ombre
Euclidean Distance from input image: 48.101875
Amazon Url: www.amazon.com/dp/B01F7PHXY8





Product Title: summer alpaca 3d pastel casual loose tops tee design

Euclidean Distance from input image: 48.118896

Amazon Url: www.amazon.com/dp/B01I80A93G



Product Title: miss chievous juniors striped peplum tank top medium shadowpeach

Euclidean Distance from input image: 48.13128

Amazon Url: www.amazon.com/dp/B0177DM70S



Product Title: red pink floral heel sleeveless shirt xl xxl

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



Product Title: moana logo adults hot v neck shirt black xxl

Euclidean Distance from input image: 48.25678

Amazon Url: www.amazon.com/dp/B01LX6H43D



Product Title: abaday multicolor cartoon cat print short sleeve longline shirt large

Euclidean Distance from input image: 48.265644

Amazon Url: www.amazon.com/dp/B01CR57YY0



Product Title: kawaii cotton pastel tops tees peach pink cactus design

Euclidean Distance from input image: 48.362583

Amazon Url: www.amazon.com/dp/B071WYLBZS



Product Title: chicago chicago 18 shirt women pink

Euclidean Distance from input image: 48.383648

Amazon Url: www.amazon.com/dp/B01GXAZTRY



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

Product Title: yichun womens tiger printed summer tshirts tops

Euclidean Distance from input image: 48.449345

Amazon Url: www.amazon.com/dp/B010NN9RX0



Product Title: nancy lopez whimsy short sleeve whiteblacklemon drop xs

Euclidean Distance from input image: 48.478893

Amazon Url: www.amazon.com/dp/B01MPX6IDX





Product Title: womens tops tees pastel peach ice cream cone print
Euclidean Distance from input image: 48.557983
Amazon Url: www.amazon.com/dp/B0734GRKZL



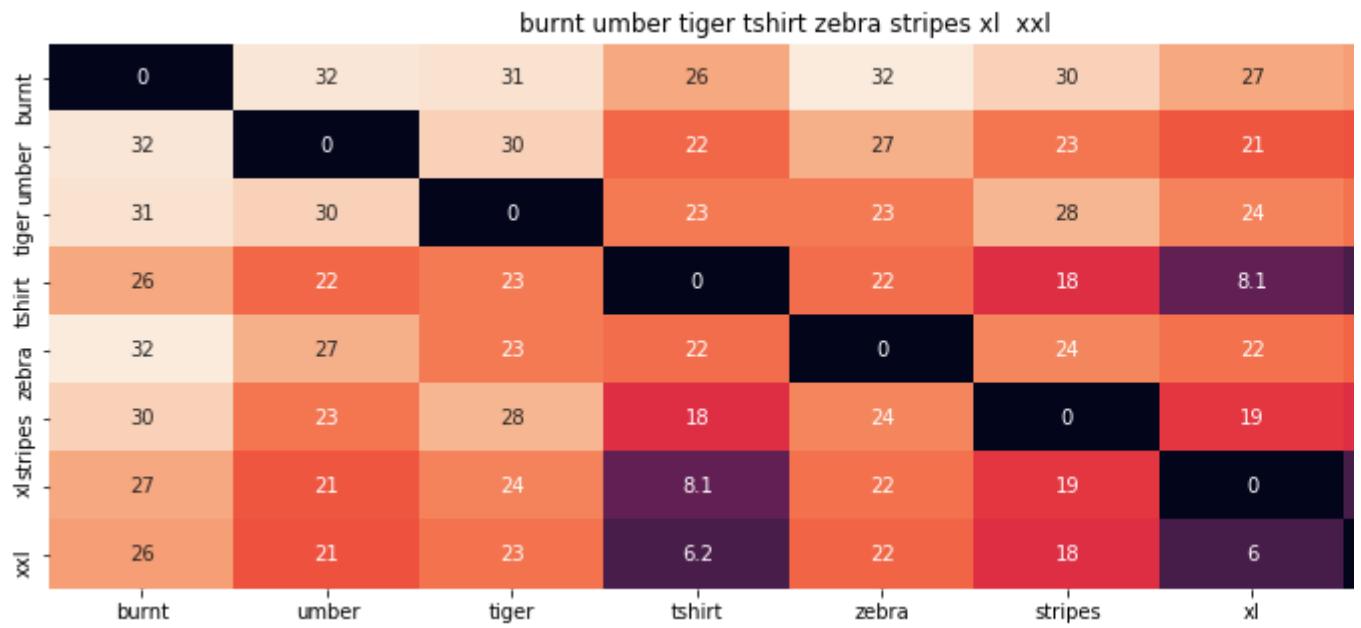
Product Title: uswomens mary j blige without tshirts shirt
Euclidean Distance from input image: 48.614395
Amazon Url: www.amazon.com/dp/B01M0XXFKK

▼ Eqaul Weightage

```
idf_w2v_brand(12566,20, 5,5,5)
```



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



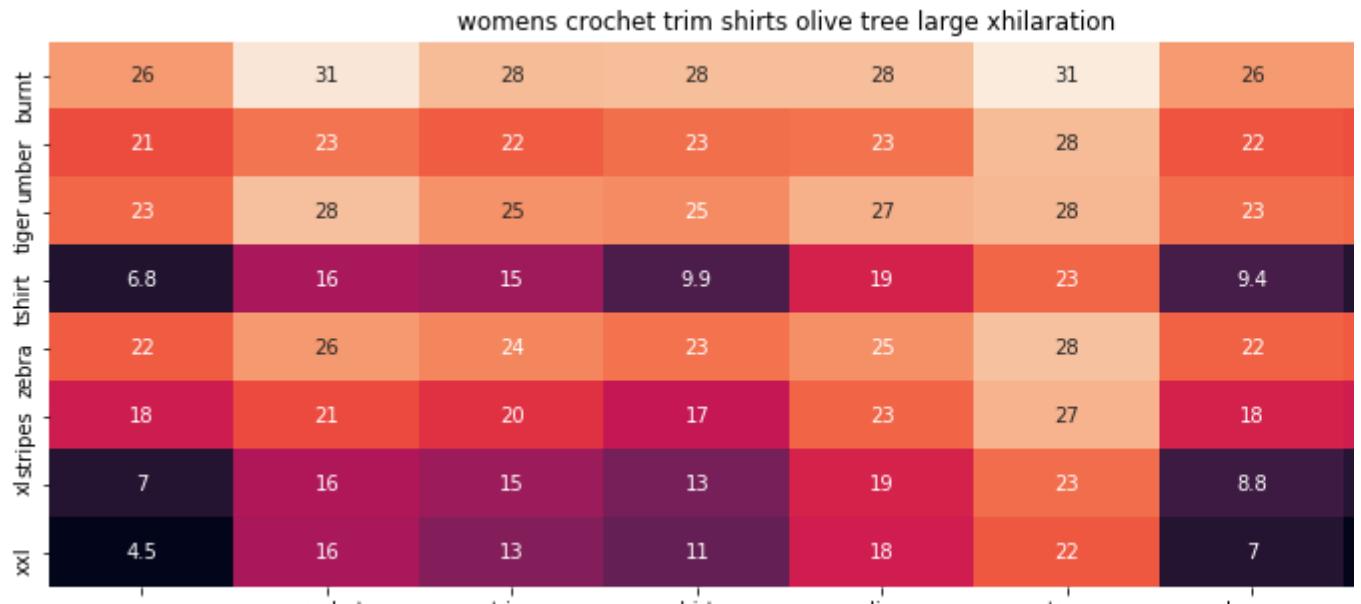
ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.00130208333333333333

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B06XBHN7J

Brand : Xhilaration

euclidean distance from input : 15.982596334062531

	buffalo	david	bitton	nipaw	logo	graphic	tank	white	combo
burnt	31	34	26	26	28	27	26	25	29
tiger	28	28	21	21	24	22	22	20	24
umber	23	32	23	23	25	24	23	22	25
tshirt	21	22	7	7	13	11	9	7.7	14
zebra	21	31	22	22	24	23	22	21	24
stripes	26	28	18	18	19	19	18	17	22
xxl	22	22	7.2	7.2	15	13	9.4	8	13
xxl	20	22	4.7	4.7	14	11	7.8	5.9	12

ACTN1 - DQ1QH4ZY00

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).



	free	people	free	malibu	thermal	henley	pullover	black	small
burnt	28	27	28	27	33	27	31	25	
umber	24	24	24	22	30	21	25	20	
tiger	26	25	26	26	31	25	26	22	
tshirt	13	13	13	13	23	9.3	14	7.6	
zebra	25	23	25	23	30	22	25	21	
stripes	21	21	21	21	29	18	21	17	

xxl	13	14	13	13	23	11	16	7.9
xxl	12	12	12	12	22	8.8	15	5.9

ASIN : B074MXY984

Brand : We The Free

euclidean distance from input : 16.9315793355306

j brand womens pinstripe shirt xs blue					
burnt	29	26	36	26	26
tiger	27	21	29	22	21
umber	27	23	35	23	23
tshirt	16	6.8	26	6	7.7
zebra	26	22	33	22	23
stripes	23	18	26	17	19
xl	16	7	27	8.4	6.9

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

j	brand	womens	pinstripe	shirt
---	-------	--------	-----------	-------

ASIN : B06XYP1X1F

Brand : J Brand Jeans

euclidean distance from input : 16.955509586267063

completely liz lange long flyaway vest 249682 turquoise						
burnt	32	30	31	26	36	28
tiger	33	24	23	22	33	24
umber	32	27	28	23	34	25
tshirt	25	15	17	8.6	26	13
xl						7

	completely	liz	lange	long	flyaway	vest	249682
xl stripes zebra	32	26	26	22	32	24	22
xxl	30	23	24	19	32	19	18
xl	25	16	17	8.5	26	13	7.2
xxl	25	15	16	6.6	26	12	4.7

ASIN : B074LTBWSW

Brand : Liz Lange

euclidean distance from input : 16.957030995927767

	bila	size	small	womens	sleeveless	blouse	red
burnt tiger	26	26	26	26	27	26	26
umber	22	21	21	21	22	21	21
burnt tiger	24	23	23	23	25	23	23
umber	13	8.3	8.5	6.8	11	7.3	7.3
burnt tiger	24	22	22	22	23	22	22

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

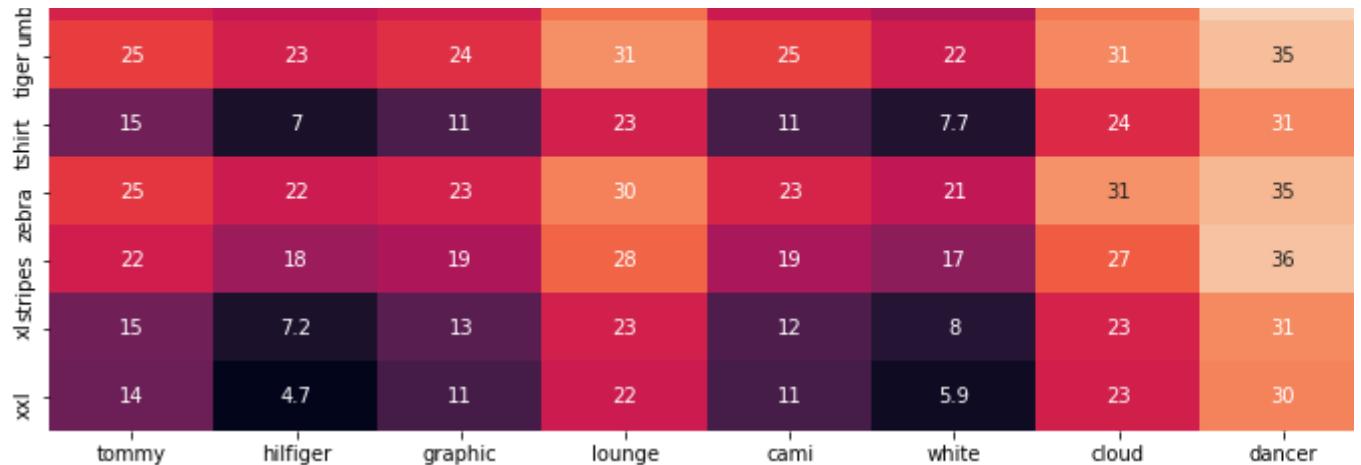
	bila	size	small	womens	sleeveless	blouse
xxl	12	7.7	8.4	7	12	8.4
xl	11	5.9	6.4	4.5	10	6.2

ASIN : B01L7ROZNC

Brand : Bila

euclidean distance from input : 17.094785603444056

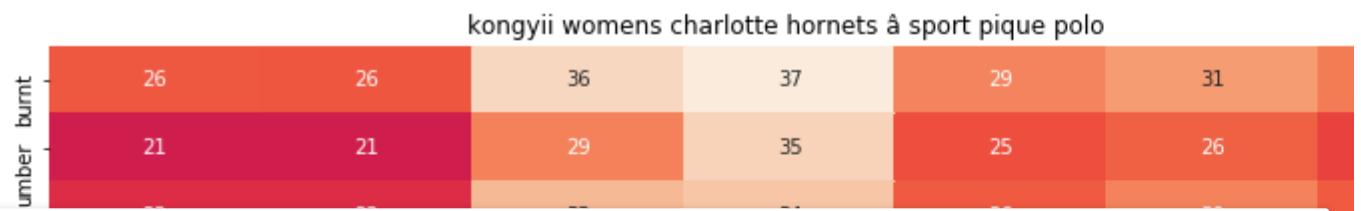
	tommy hilfiger	graphic	lounge	cami	white	cloud	dancer	xl
burnt	29	26	27	33	28	25	34	39
umber	23	21	22	29	22	20	29	37



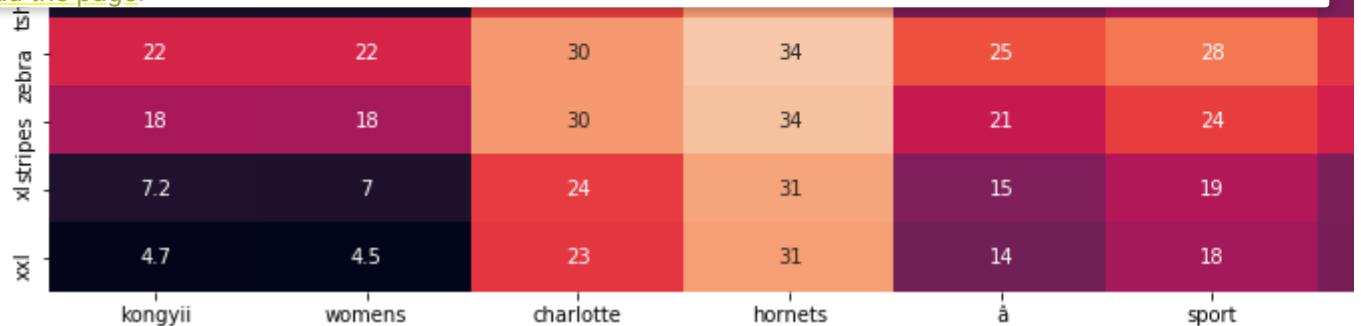
ASIN : B01BMSFYW2

Brand : igertommy hilf

euclidean distance from input : 17.28112604134519



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B01FJVZST2

Brand : KONGYII

euclidean distance from input : 17.551688170989628

	32	27	26	26	31	30	32	32	28	26	26
burnt umber	28	23	21	22	27	24	26	28	22	21	20
tiger tshirt	29	24	23	23	28	27	28	30	25	23	23
zebra tshirt	20	11	7	7.3	19	19	18	20	11	7.3	8.1
zebra stripes	28	23	22	22	28	25	28	29	24	22	21
zebra stripes	25	20	18	17	24	24	26	26	18	18	17
zebra stripes	20	11	7.2	8.3	19	20	19	20	12	7.5	9.1
xxl	19	9.4	4.7	6.6	18	19	18	19	11	5	7.4
	boundaries	juniors	34	sleeve	space	dye	hi	lo	knit	top	pink

ASIN : B01EXXFS4M

Brand : No Boundaries

euclidean distance from input : 17.564531453691437

stanzino womens long sleeve graphic print plus size top fuchsia xl

	23	23	23	23	24	23	23	23	23	29	2
tiger tshirt	7	6.8	8.6	7.3	11	9.7	8.4	8.3	7.3	21	4
zebra tshirt	22	22	22	22	23	22	23	22	22	26	
zebra stripes	18	18	19	17	19	19	18	18	18	24	
zebra stripes	7.2	7	8.5	8.3	13	10	8.1	7.7	7.5	20	
xxl	4.7	4.5	6.6	6.6	11	8.7	6.2	5.9	5	20	
	stanzino	womens	long	sleeve	graphic	print	plus	size	top	fuchsia	

ASIN : B00DP4VHWI

Brand : Stanzino

euclidean distance from input : 17.585476724863327

1state womens medium chambray crochet solid blouse blue

	1state	womens	medium	chambray	crochet	solid	blouse
burnt	26	26	26	30	31	27	26
tiger	21	21	22	23	23	23	21
umber	23	23	24	28	28	25	23
tshirt	7	6.8	9.7	16	16	12	7.3
zebra	22	22	23	25	26	24	22
stripes	18	18	19	21	21	20	18
xl	7.2	7	9.4	17	16	12	8.4
xxl	4.7	4.5	8	16	16	11	6.2

ASIN : B074MK6LV2

Brand : 1.State

euclidean distance from input : 17.59606810466476

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

byoung womens henya womens light blue shirt size 40l light blue

	26	26	26	26	27	26	26	26	26	27
burnt	26	26	26	26	27	26	26	26	26	27
tiger	21	21	21	21	22	20	22	21	21	22
umber	23	23	23	23	25	23	23	23	23	25
tshirt	7	6.8	7	6.8	13	8	6	8.3	7	13
zebra	22	22	22	22	24	21	22	22	22	24
stripes	18	18	18	18	19	16	17	18	18	19
xl	7.2	7	7.2	7	13	8.7	8.4	7.7	7.2	13
xxl	4.7	4.5	4.7	4.5	12	7	6.3	5.9	4.7	12

ASIN : B06Y41MRCH

Brand : Byoung

euclidean distance from input : 17.624772135657583

=====

	usstore	women	stripes	oversized	beach	shirt	long	sleeve	casual	blouse	tee	tops	size
xxl	26	26	30	29	28	26	26	26	27	26	27	26	1
tiger	21	22	23	24	26	22	22	22	22	21	22	22	umber
zebra	23	23	28	26	27	23	23	23	24	23	24	24	stripes
tskirt	7	9	18	15	16	6	8.6	7.3	9.9	7.3	11	9.6	burnt
xxl	22	23	24	23	25	22	22	22	23	22	23	23	burnt
xxl	18	19	0	21	23	17	19	17	19	18	20	18	burnt
xxl	7.2	9.3	19	16	17	8.4	8.5	8.3	11	8.4	12	9.8	burnt
xxl	4.7	7.5	18	15	15	6.3	6.6	6.6	8.9	6.2	10	8.1	burnt

ASIN : B01DNNI1RO

Brand : Usstore

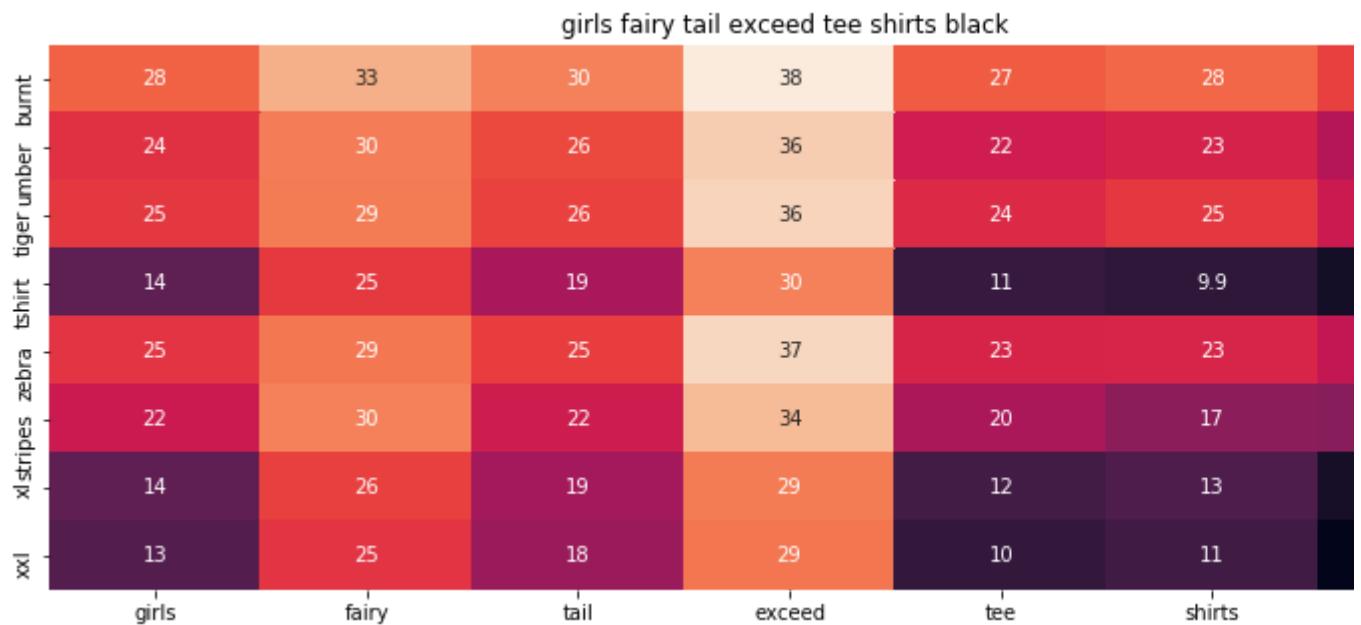
euclidean distance from input : 17.65560150170576

	salvatore	ferragamo	geometric	print	silk	top	6
xxl	26	26	32	26	27	26	1
tiger	21	21	25	22	21	21	umber
zebra	23	23	32	23	24	23	stripes
tskirt	7	7	22	9.7	12	7.3	burnt
xxl	22	22	27	22	23	22	burnt
xxl	18	18	24	19	19	18	burnt
xxl	7.2	7.2	22	10	13	7.5	burnt
xxl	4.7	4.7	21	8.7	11	5	burnt

ASIN : B0756JTS1F

Brand : Salvatore Ferragamo

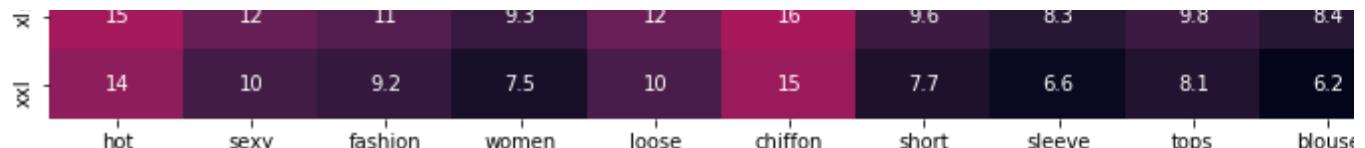
euclidean distance from input : 17.681495666503906



ASIN : B01L9F153U

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

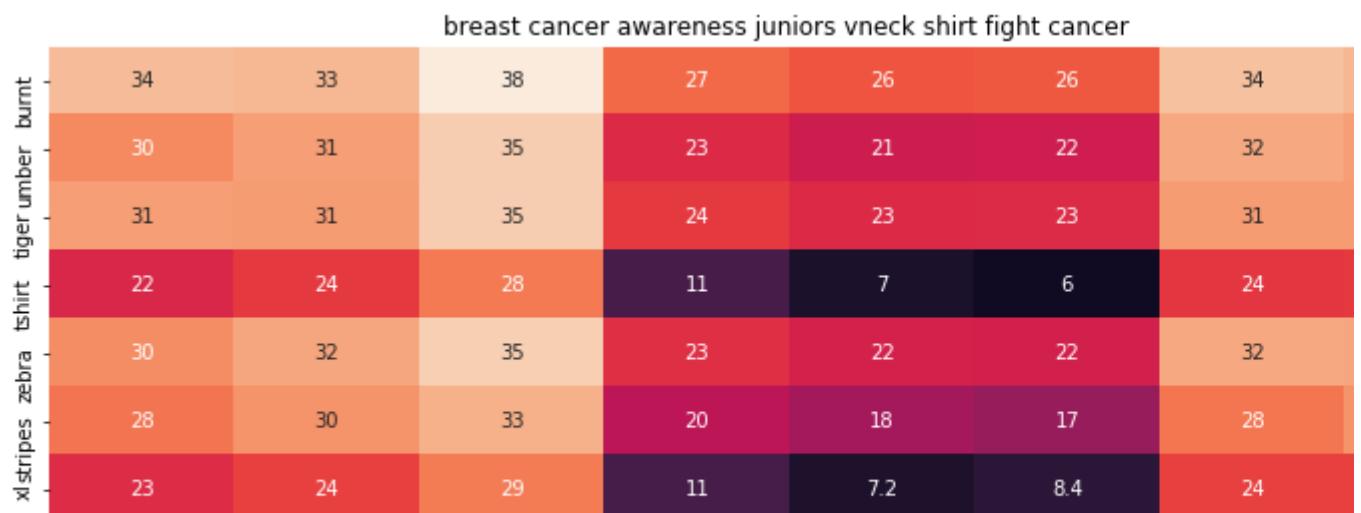




ASIN : B00JMAASRO

Brand : Wotefusi

euclidean distance from input : 17.70576048747726

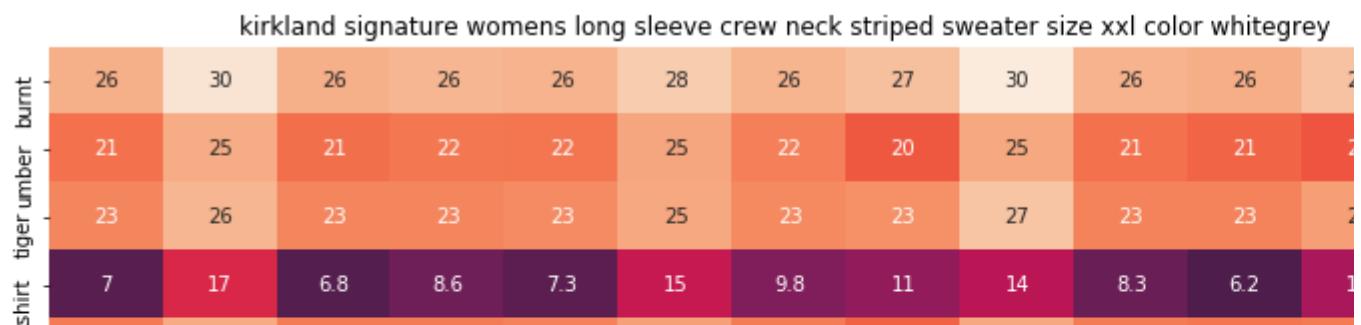


Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B016CU40IY

Brand : Juiceclouds

euclidean distance from input : 17.709572641611373



	22	25	22	22	22	25	22	20	25	22	22	2
xxl	4.7	16	4.5	6.6	6.6	14	8.7	10	15	5.9	0	1
size	7.2	17	7	8.5	8.3	16	10	12	17	7.7	6	1
color	22	25	22	22	22	25	22	20	25	22	22	2
neck	18	21	18	19	17	22	18	16	20	18	18	1
long	21	26	21	21	21	22	22	22	22	22	22	2
sleeve	23	25	23	23	23	23	23	23	23	23	23	2
roundedv	7.8	19	6.8	7	7	9.8	8.6	8.6	9	9	9	2
womens	22	26	22	22	22	22	22	22	22	22	22	2
signature	26	31	26	26	26	26	26	26	26	26	26	2
kirkland	21	26	21	21	21	22	22	22	22	22	22	2
brand	23	25	23	23	23	23	23	23	23	23	23	2
xxl	7.8	19	6.8	7	7	9.8	8.6	8.6	9	9	9	2
size	22	26	22	22	22	22	22	22	22	22	22	2

ASIN : B06XTPC3FP

Brand : Kirkland Signature

euclidean distance from input : 17.753979238115267

	eileen fisher womens roundedv neck long tank gray l						
xxl	26	31	26	26	26	26	26
size	21	26	21	21	22	22	22
color	23	25	23	23	23	23	23
neck	7.8	19	6.8	7	9.8	8.6	9
long	22	26	22	22	22	22	22
womens	26	31	26	26	26	26	26
roundedv	21	26	21	21	22	22	22
brand	23	25	23	23	23	23	23
xxl	7.8	19	6.8	7	9.8	8.6	9
size	22	26	22	22	22	22	22

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	eileen	fisher	womens	roundedv	neck	long	tank
xxl	8	19	7	7.2	10	8.5	9.4
size	5.8	19	4.5	4.7	8.7	6.6	7.8
color	22	26	22	22	22	22	22

ASIN : B01N2JSRDM

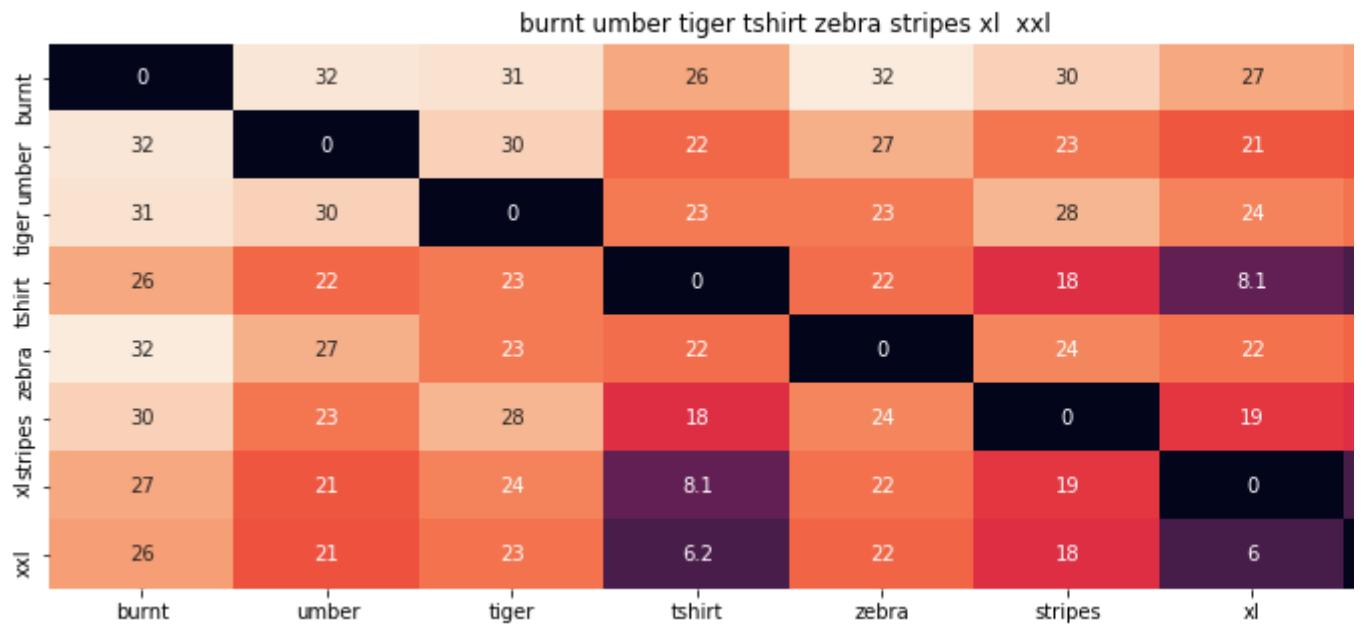
Brand : Eileen Fisher

euclidean distance from input : 17.801957957190787

- When Image feature has maximum weightage than 'Title' and '

```
idf_w2v_brand(12566,20, 5,5,50)
```





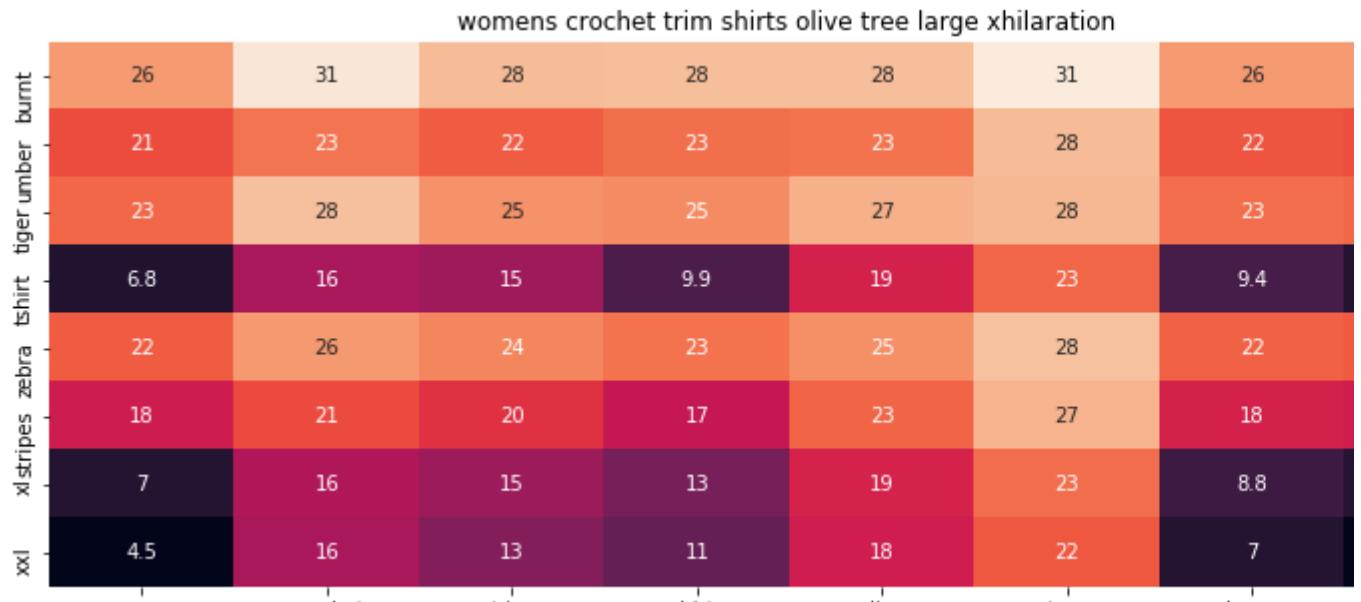
ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.00032552083333333333

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



WOMENS

CROCHET

WIM

SHIRTS

VNECK

TEE

TSHIRT

ASIN : B06XBHN7J

Brand : Xhilaration

euclidean distance from input : 31.88651212062501

	breast	cancer	awareness	juniors	vneck	shirt	fight
breast	34	33	38	27	26	26	34
cancer	30	31	35	23	21	22	32
awareness	31	31	35	24	23	23	31
juniors	22	24	28	11	7	6	24
vneck	30	32	35	23	22	22	32
shirt	28	30	33	20	18	17	28
fight	23	24	29	11	7.2	8.4	24
breast	22	24	29	9.4	4.7	6.3	24

ACTUAL DOGMA

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	free	people	free	malibu	thermal	henley	pullover	black	small
burnt	28	27	28	27	33	27	31	25	
umber	24	24	24	22	30	21	25	20	
tiger	26	25	26	26	31	25	26	22	
tshirt	13	13	13	13	23	9.3	14	7.6	
zebra	25	23	25	23	30	22	25	21	
ripenes	21	21	21	21	29	18	21	17	

xxl	13	14	13	13	23	11	16	7.9
xxl	12	12	12	12	22	8.8	15	5.9

ASIN : B074MXY984

Brand : We The Free

euclidean distance from input : 33.75603326161703

completely liz lange long flyaway vest 249682 turquoise								
burnt	32	30	31	26	36	28	26	
tiger	33	24	23	22	33	24	21	
umber	32	27	28	23	34	25	23	
tshirt	25	15	17	8.6	26	13	7	
zebra	32	26	26	22	32	24	22	
stripes	30	23	24	19	32	19	18	
xl	25	16	17	8.5	26	13	7.2	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).

completely	liz	lange	long	long	flyaway	vest	249682
ASIN : B074LTBWSW							
Brand : Liz Lange							
euclidean distance from input : 33.762396176716315							

buffalo david bitton nipaw logo graphic tank white combo xxl								
burnt	31	34	26	26	28	27	26	25
tiger	28	28	21	21	24	22	22	20
umber	23	32	23	23	25	24	23	22
tshirt	21	22	7	7	13	11	9	7.7

size	zebra	31	22	22	24	23	22	21	24
stripes	21	31	22	22	24	23	22	21	24
xl	26	28	18	18	19	19	18	17	22
xxl	22	22	7.2	7.2	15	13	9.4	8	13
xxxl	20	22	4.7	4.7	14	11	7.8	5.9	12
	buffalo	david	bitton	nipaw	logo	graphic	tank	white	combo

ASIN : B018H5AZXQ

Brand : Buffalo

euclidean distance from input : 33.82252559667889

j brand womens pinstripe shirt xs blue									
size	burnt	29	26	36	26	26	26	26	26
burnt	29	26	36	26	26	26	26	26	26
tiger	27	21	29	22	21	21	21	21	21
umber	27	23	35	23	23	23	23	23	23
tshirt	16	6.8	26	6	7.7	7.7	7.7	7.7	7.7
xxl	26	22	33	22	22	22	22	22	22
	j	brand	womens	pinstripe	shirt				

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

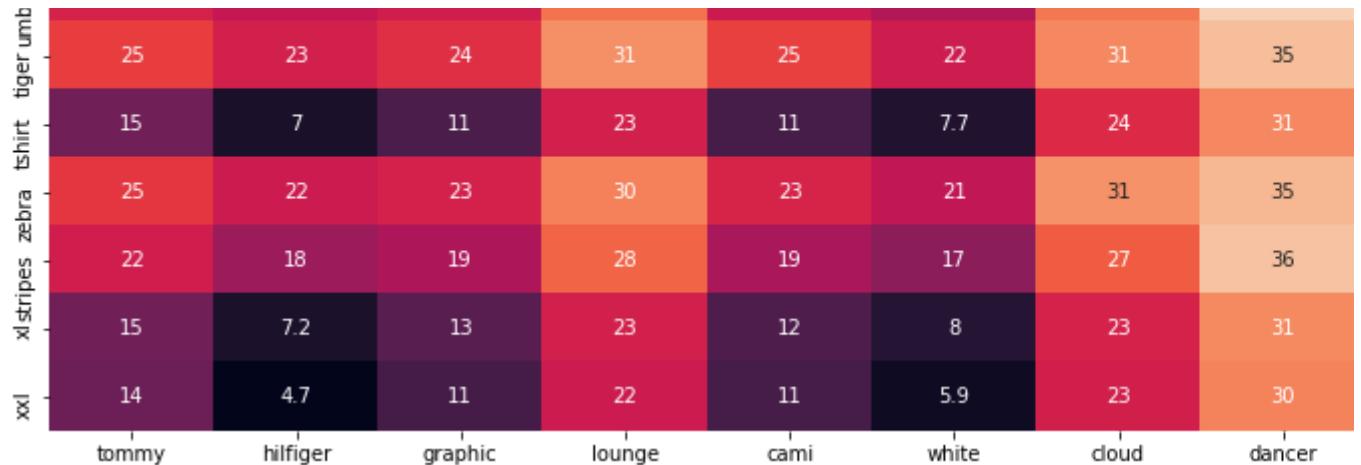
size	burnt	16	7	27	8.4	6.9	6.9	6.9	6.9
burnt	16	7	27	8.4	6.9	6.9	6.9	6.9	6.9
xxl	15	4.5	27	6.3	5.8	5.8	5.8	5.8	5.8
	j	brand	womens	pinstripe	shirt				

ASIN : B06XYP1X1F

Brand : J Brand Jeans

euclidean distance from input : 33.93180393058369

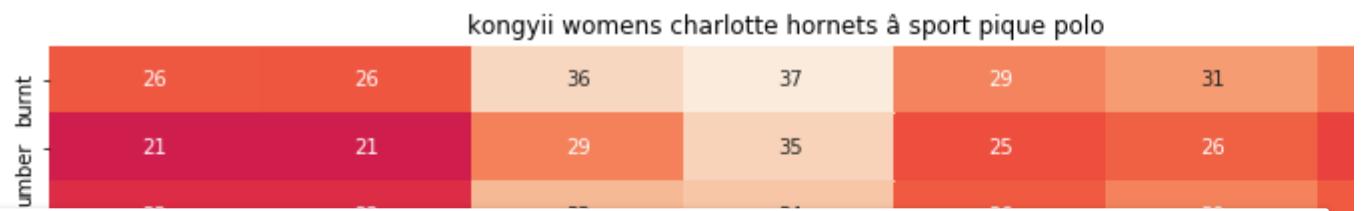
tommy hilfiger graphic lounge cami white cloud dancer xlarge									
size	burnt	29	26	27	33	28	25	34	39
burnt	29	26	27	33	28	25	34	39	37
xxl	23	21	22	29	22	20	29	37	
	j	brand	womens	pinstripe	shirt				



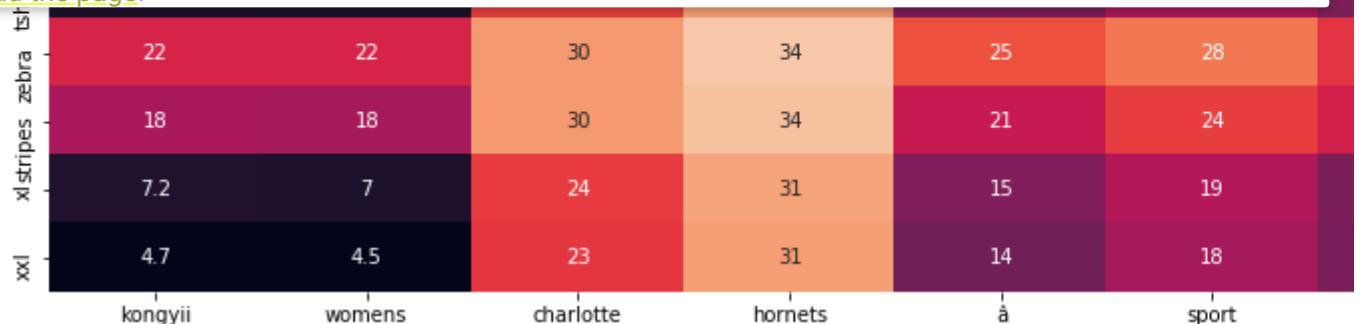
ASIN : B01BMSFYW2

Brand : igertommy hilf

euclidean distance from input : 34.12049869376729



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B01FJVZST2

Brand : KONGYII

euclidean distance from input : 34.933592615664075

	26	26	26	26	27	26	
burnt	26	26	26	26	27	26	
tiger umber	22	21	21	21	22	21	
tiger umber	24	23	23	23	25	23	
tshirt	13	8.3	8.5	6.8	11	7.3	
zebra	24	22	22	22	23	22	
zebra	21	18	19	18	18	18	
xlstripes	12	7.7	8.4	7	12	8.4	
xxl	11	5.9	6.4	4.5	10	6.2	
	bila	size	small	womens	sleeveless	blouse	

ASIN : B01L7ROZNC

Brand : Bila

euclidean distance from input : 35.05102285756674

=====

girls fairy tail exceed tee shirts black

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	25	29	26	36	24	25	
tiger umber	25	29	26	36	24	25	
tshirt	14	25	19	30	11	9.9	
zebra	25	29	25	37	23	23	
xlstripes	22	30	22	34	20	17	
xlstripes	14	26	19	29	12	13	
xxl	13	25	18	29	10	11	
	girls	fairy	tail	exceed	tee	shirts	

ASIN : B01L9F153U

Brand : ATYPEMX

euclidean distance from input : 35.61336702718344

=====

	boundaries	juniors	34	sleeve	space	dye	hi	lo	knit	top	medium
xxl	32	27	26	26	31	30	32	32	28	26	26
xl	28	23	21	22	27	24	26	28	22	21	20
l	29	24	23	23	28	27	28	30	25	23	23
m	20	11	7	7.3	19	19	18	20	11	7.3	8.1
s	28	23	22	22	28	25	28	29	24	22	21
xs	25	20	18	17	24	24	26	26	18	18	17
xs	20	11	7.2	8.3	19	20	19	20	12	7.5	9.1
xxs	19	9.4	4.7	6.6	18	19	18	19	11	5	7.4
	boundaries	juniors	34	sleeve	space	dye	hi	lo	knit	top	medium

ASIN : B01EXXFS4M

Brand : No Boundaries

euclidean distance from input : 35.658698527078464

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	salvatore	ferragamo	geometric	print	silk	top
xxl	26	26	32	26	27	26
xl	21	21	25	22	21	21
l	23	23	32	23	24	23
m	7	7	22	9.7	12	7.3
s	22	22	27	22	23	22
xs	18	18	24	19	19	18
xs	7.2	7.2	22	10	13	7.5
xxs	4.7	4.7	21	8.7	11	5
	salvatore	ferragamo	geometric	print	silk	top

ASIN : B0756JTS1F

Brand : Salvatore Ferragamo

euclidean distance from input : 35.8234561920166

	byoung	womens	henya	womens	light	blue	shirt	size	40l	light
xxl	26	26	26	26	27	26	26	26	26	27
tiger	21	21	21	21	22	20	22	21	21	22
umbre	23	23	23	23	25	23	23	23	23	25
tshirt	7	6.8	7	6.8	13	8	6	8.3	7	13
zebra	22	22	22	22	24	21	22	22	22	24
stripes	18	18	18	18	19	16	17	18	18	19
xl	7.2	7	7.2	7	13	8.7	8.4	7.7	7.2	13
xxl	4.7	4.5	4.7	4.5	12	7	6.3	5.9	4.7	12

ASIN : B06Y41MRCH

Brand : Byoung

euclidean distance from input : 35.8665211996045

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

	1state	womens	medium	chambray	crochet	solid	blouse
xxl	26	26	26	30	31	27	26
tiger	21	21	22	23	23	23	21
umbre	23	23	24	28	28	25	23
tshirt	7	6.8	9.7	16	16	12	7.3
zebra	22	22	23	25	26	24	22
stripes	18	18	19	21	21	20	18
xl	7.2	7	9.4	17	16	12	8.4
xxl	4.7	4.5	8	16	16	11	6.2

ASIN : B074MK6LV2

Brand : 1.State

euclidean distance from input : 35.93611775451905

		sexy sheer mesh print long sleeves bodysuit					
		sexy	sheer	mesh	print	long	sleeves
burnt	27	28	29	26	26	29	
tiger	22	23	23	22	22	24	
umber	24	25	25	23	23	27	
tshirt	11	13	14	9.7	8.6	14	
zebra	23	25	23	22	22	26	
stripes	20	21	19	19	19	19	
xl	12	13	14	10	8.5	15	
xxl	10	12	13	8.7	6.6	14	

ASIN : B074Z5C98D

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

stanzino womens long sleeve graphic print plus size top fuchsia xl										
	26	26	26	26	27	26	27	26	26	32
burnt	26	26	26	26	27	26	27	26	26	32
tiger	21	21	22	22	22	22	21	21	21	24
umber	23	23	23	23	24	23	23	23	23	29
tshirt	7	6.8	8.6	7.3	11	9.7	8.4	8.3	7.3	21
zebra	22	22	22	22	23	22	23	22	22	26
stripes	18	18	19	17	19	19	18	18	18	24
xxl	12	7	8.5	8.2	12	10	8.1	7.7	7.5	20

xxl	7.2	7	8.5	8.3	13	10	8.1	7.7	7.5	20
xxl	4.7	4.5	6.6	6.6	11	8.7	6.2	5.9	5	20

ASIN : B00DP4VHWI

Brand : Stanzino

euclidean distance from input : 36.17867763255047

	maven west striped sleeveless lace peplum peasant blouse yellow large									
xxl	41	29	27	27	27	31	28	26	27	
burnt tiger	37	25	20	22	21	22	26	21	20	
umber	37	27	23	25	24	28	26	23	24	
tshirt	30	17	11	11	10	17	17	7.3	13	
zebra	36	25	20	23	23	25	26	22	21	
stripes	35	23	16	18	17	21	24	18	17	
xxl	31	17	12	12	11	17	18	8.4	13	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

ASIN : B01M8GB3AL

Brand : Maven West

euclidean distance from input : 36.21196498876873

	hot sexy fashion women loose chiffon short sleeve tops blouse shirt									
xxl	27	27	27	26	27	29	27	26	26	26
burnt tiger	24	22	22	22	23	21	22	22	22	21
umber	25	24	24	23	24	27	23	23	24	23
tshirt	15	11	10	9	12	15	9.3	7.3	9.6	7.3

zebra	25	23	23	23	22	24	23	22	23	22
stripes	21	20	19	19	20	20	19	17	18	18
xl	15	12	11	9.3	12	16	9.6	8.3	9.8	8.4
xxl	14	10	9.2	7.5	10	15	7.7	6.6	8.1	6.2
	hot	sexy	fashion	women	loose	chiffon	short	sleeve	tops	blouse

ASIN : B00JMAASRO

Brand : Wotefusi

euclidean distance from input : 36.23066985819744

womens ultimate scoop tee fresh white xl merona							
burnt	26	32	27	27	28	25	27
tiger	21	29	23	22	25	20	21
umber	23	29	25	24	26	22	24
tshirt	6.8	19	13	11	17	7.7	8.1
merona	22	28	24	23	26	21	22
	womens	ultimate	scoop	tee	fresh	white	xl

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

xs	7	19	14	12	17	8	0
xl	4.5	18	12	10	16	5.9	6
	womens	ultimate	scoop	tee	fresh	white	xl

ASIN : B01G7XE50E

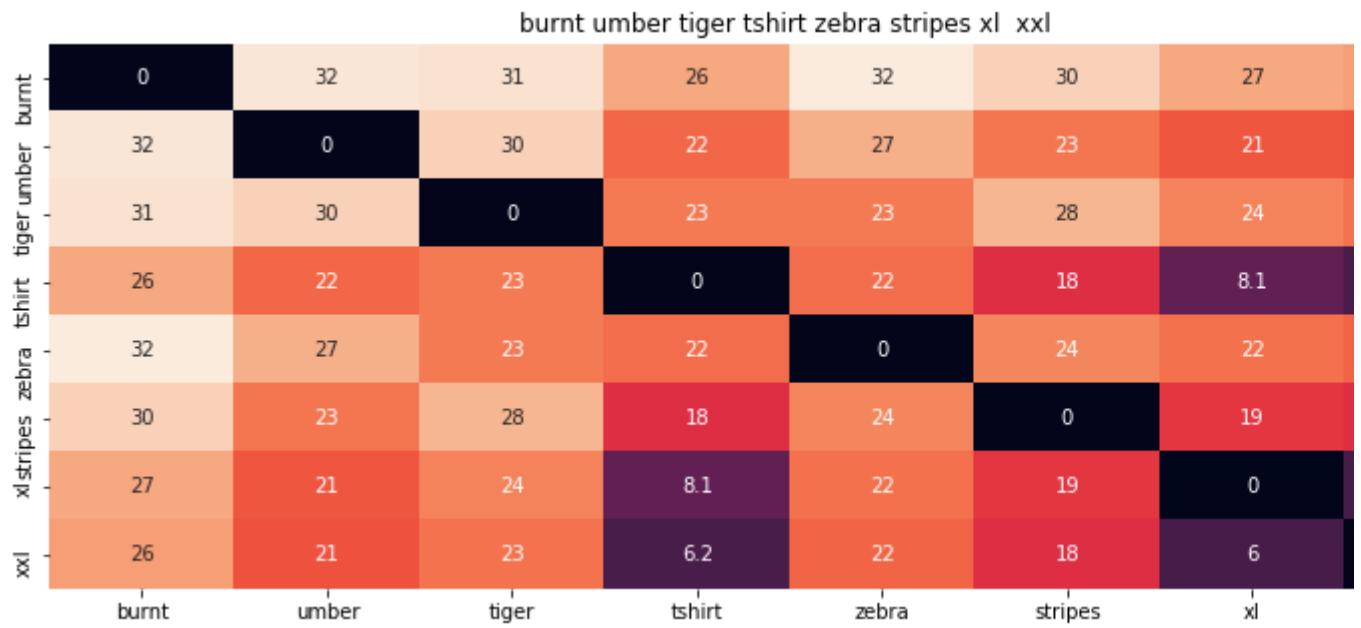
Brand : Merona

euclidean distance from input : 36.25837619517254

- When 'Brand & Color' has maximum weightage than 'Image features'

```
idf_w2v_brand(12566,20, 1,10,1)
```





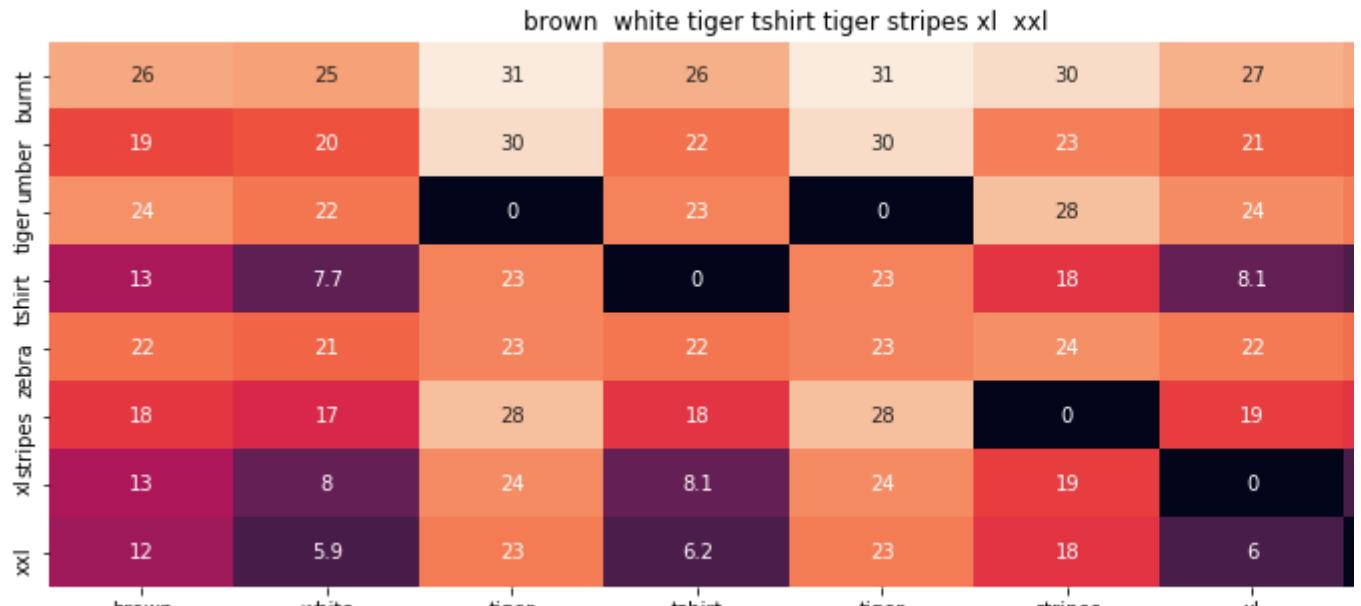
ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.00032552083333333333

=====

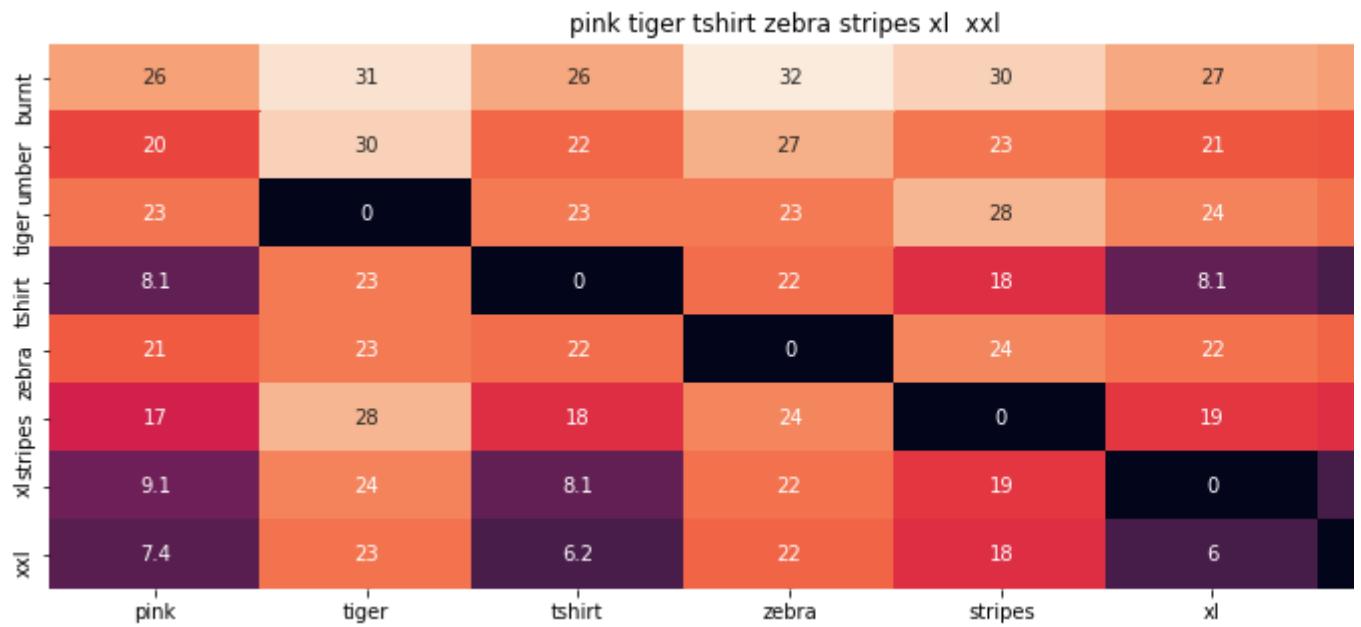
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B00JXQCWT0

Brand : Si Row

euclidean distance from input : 5.026570280392964



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

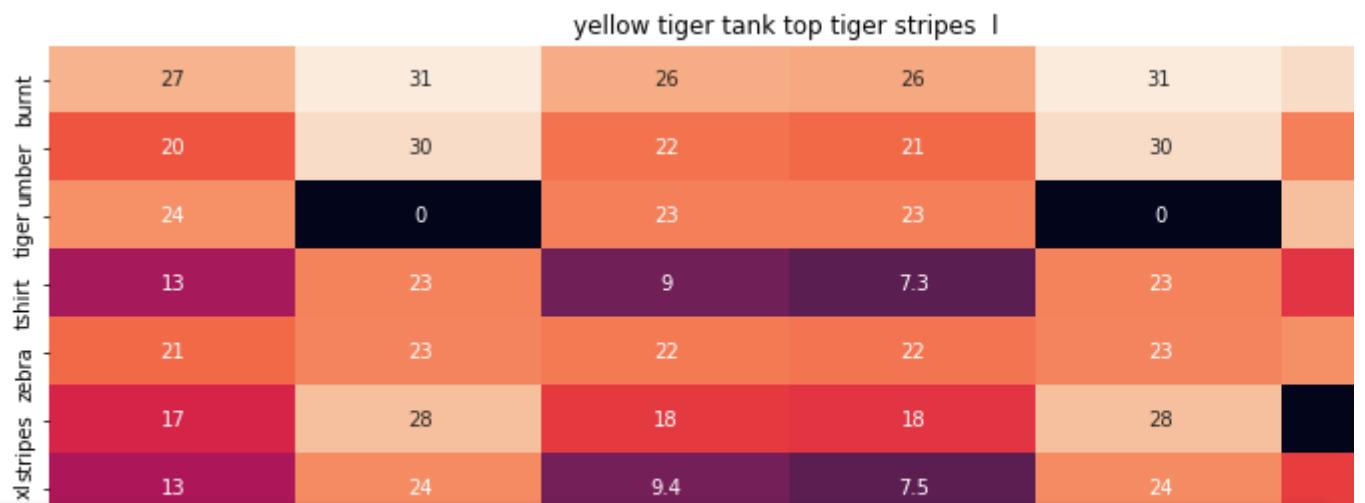




ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from input : 6.017716606759175



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQAUWA

Brand : Si Row

euclidean distance from input : 6.109212676985209



zebra	22	26	24	23	25	28	22
stripes	18	21	20	17	23	27	18
xl	7	16	15	13	19	23	8.8
xxl	4.5	16	13	11	18	22	7

ASIN : B06XBHM7J

Brand : Xhilaration

euclidean distance from input : 6.1169693873388455

	grey	white	tiger	tank	top	tiger	stripes	xl	xxl
burnt	26	25	31	26	26	31	30	27	
umber	21	20	30	22	21	30	23	21	
tiger	23	22	0	23	23	0	28	24	
tshirt	7	7.7	23	9	7.3	23	18	8.1	
	22	21	23	22	22	23	24	22	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

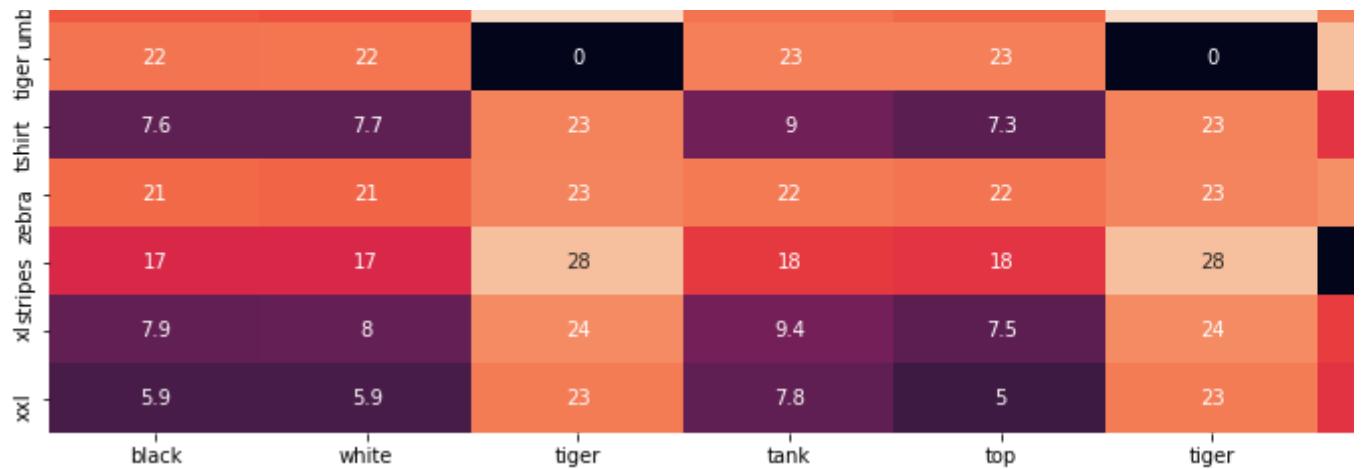
xxl	7.2	8	24	9.4	7.5	24	19	0
xl	4.7	5.9	23	7.8	5	23	18	6

ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from input : 6.119070093138163

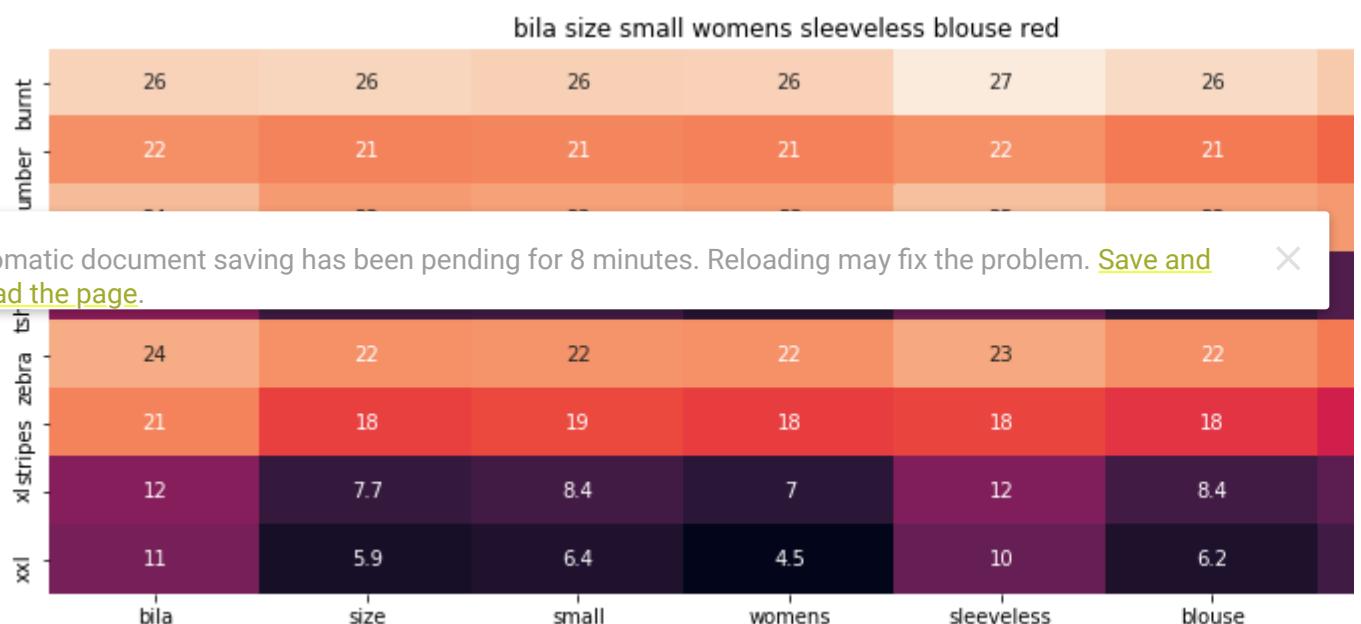
	black	white	tiger	tank	top	tiger	stripes	l
burnt	25	25	31	26	26	31		
umber	20	20	30	22	21	30		



ASIN : B00JXQA094

Brand : Si Row

euclidean distance from input : 6.25537292192692



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B01L7ROZNC

Brand : Bila

euclidean distance from input : 6.258009892106743

	31	34	26	26	28	27	26	25	29
burnt	31	34	26	26	28	27	26	25	29
tiger umber	28	28	21	21	24	22	22	20	24
tiger	23	32	23	23	25	24	23	22	25
tshirt	21	22	7	7	13	11	9	7.7	14
zebra	21	31	22	22	24	23	22	21	24
stripes	26	28	18	18	19	19	18	17	22
xl	22	22	7.2	7.2	15	13	9.4	8	13
xxl	20	22	4.7	4.7	14	11	7.8	5.9	12
	buffalo	david	bitton	nipaw	logo	graphic	tank	white	combo

ASIN : B018H5AZXQ

Brand : Buffalo

euclidean distance from input : 6.272295117980529

hip latter crochet back womens small hilow blouse brown

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	27	27	28	23	23	23	23	23
tiger umber	27	27	28	23	23	23	23	23
tshirt	17	18	16	10	6.8	8.5	7	7.3
zebra	26	27	26	23	22	22	22	22
stripes	23	25	21	19	18	19	18	18
xl	17	18	16	11	7	8.4	7.2	8.4
xxl	16	17	16	8.8	4.5	6.4	4.7	6.2
	hip	latter	crochet	back	womens	small	hilow	blouse

ASIN : B074MJN1K9

Brand : Hip

euclidean distance from input : 6.350509496986722

acquaas womens long sleeve stripe pocket fashion tshirt picture

	26	26	26	26	29	27	27	26
burnt	26	26	26	26	29	27	27	26
tiger	21	21	22	22	22	25	22	22
umber	23	23	23	23	26	26	24	23
tshirt	7	6.8	8.6	7.3	14	14	10	0
zebra	22	22	22	22	24	26	23	22
stripes	18	18	19	17	15	22	19	18
xl	7.2	7	8.5	8.3	15	15	11	8.1
xxl	4.7	4.5	6.6	6.6	14	14	9.2	6.2
	acquaas	womens	long	sleeve	stripe	pocket	fashion	tshirt

ASIN : B06XK2ZRFH

Brand : Acquaas

euclidean distance from input : 6.354364757957946

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

completely liz lange long flyaway vest 249682 turquoise

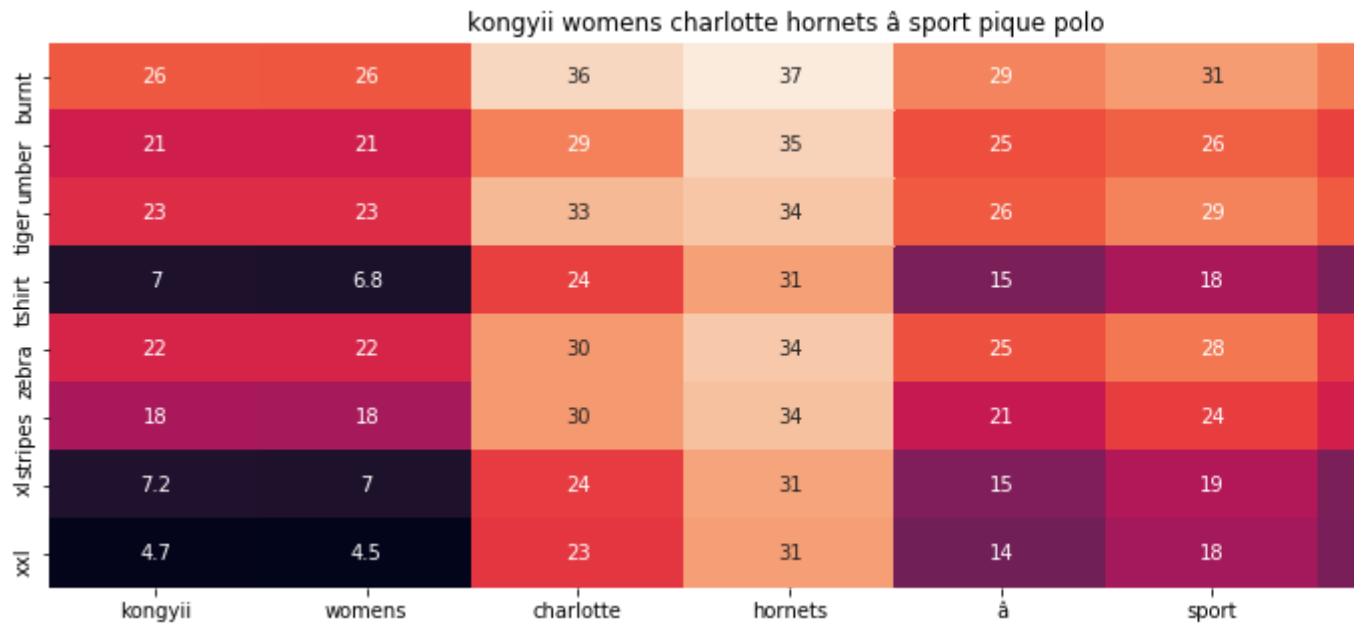
	32	30	31	26	36	28	26
burnt	32	30	31	26	36	28	26
tiger	33	24	23	22	33	24	21
umber	32	27	28	23	34	25	23
tshirt	25	15	17	8.6	26	13	7
zebra	32	26	26	22	32	24	22
stripes	30	23	24	19	32	19	18
xl	25	16	17	8.5	26	13	7.2
xxl	25	15	16	6.6	26	12	4.7
	completely	liz	lange	long	flyaway	vest	249682

ASIN : B074LTBWSW

Brand : Liz Lange

euclidean distance from input : 6.3605781402253

=====



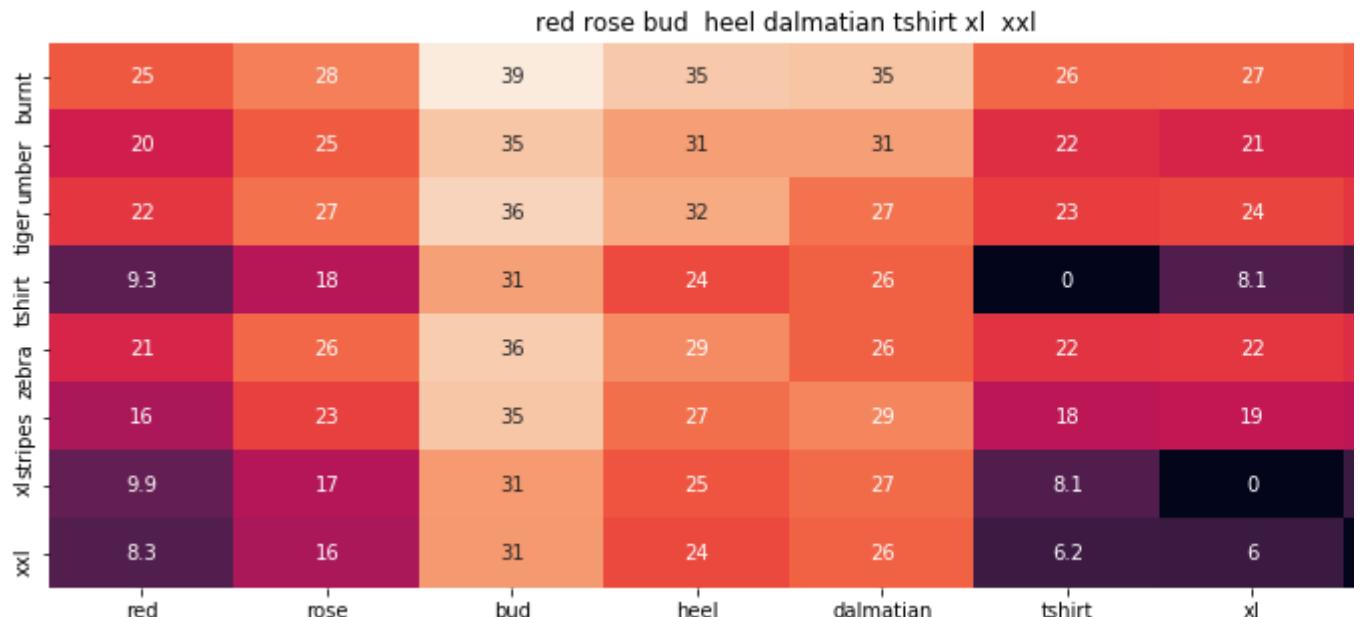
ASIN : B01FJVZST2

Brand : KONGYII

euclidean distance from input : 6.372235398889274

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B00JXQABB0

Brand : Si Row

euclidean distance from input : 6.380368153555338

stanzino womens long sleeve graphic print plus size top fuchsia xl										
burnt	26	26	26	26	27	26	27	26	26	32
tiger	21	21	22	22	22	22	21	21	21	24
umber	23	23	23	23	24	23	23	23	23	29
tshirt	7	6.8	8.6	7.3	11	9.7	8.4	8.3	7.3	21
zebra	22	22	22	22	23	22	23	22	22	26
stripes	18	18	19	17	19	19	18	18	18	24
xxl	7.2	7	8.5	8.3	13	10	8.1	7.7	7.5	20
	4.7	4.5	6.6	6.6	11	8.7	6.2	5.9	5	20
	stanzino	womens	long	sleeve	graphic	print	plus	size	top	fuchsia

ASIN : B00DP4VHWI

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

1state womens medium chambray crochet solid blouse blue							
burnt	26	26	26	30	31	27	26
tiger	21	21	22	23	23	23	21
umber	23	23	24	28	28	25	23
tshirt	7	6.8	9.7	16	16	12	7.3
zebra	22	22	23	25	26	24	22
stripes	18	18	19	21	21	20	18
xxl	7.2	7	8.4	17	16	12	8.4

xl	7.2	7	9.4	17	16	12	8.4
xxl	4.7	4.5	8	16	16	11	6.2
	1state	womens	medium	chambray	crochet	solid	blouse

ASIN : B074MK6LV2

Brand : 1.State

euclidean distance from input : 6.383330525359205

	girls fairy tail exceed tee shirts black					
xxl	28	33	30	38	27	28
xl	24	30	26	36	22	23
l	25	29	26	36	24	25
m	14	25	19	30	11	9.9
s	25	29	25	37	23	23
xs	22	30	22	34	20	17
xxs	14	26	19	29	12	13

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).

ASIN : B01L9F153U

Brand : ATYPEMX

euclidean distance from input : 6.407722097039909

	hot sexy fashion women loose chiffon short sleeve tops blouse shirt									
xxl	27	27	27	26	27	29	27	26	26	26
xl	24	22	22	22	23	21	22	22	22	21
l	25	24	24	23	24	27	23	23	24	23
m	15	11	10	9	12	15	9.3	7.3	9.6	7.3

zebra	25	23	23	23	22	24	23	22	23	22
stripes	21	20	19	19	20	20	19	17	18	18
xxl	15	12	11	9.3	12	16	9.6	8.3	9.8	8.4
	hot	sexy	fashion	women	loose	chiffon	short	sleeve	tops	blouse

ASIN : B00JMAASRO

Brand : Wotefusi

euclidean distance from input : 6.4107535892731855

	breast	cancer	awareness	juniors	vneck	shirt	fight
burnt	34	33	38	27	26	26	34
tiger	30	31	35	23	21	22	32
umber	31	31	35	24	23	23	31
tshirt	22	24	28	11	7	6	24
bra	30	32	35	23	22	22	32
	breast	cancer	awareness	juniors	vneck	shirt	fight

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

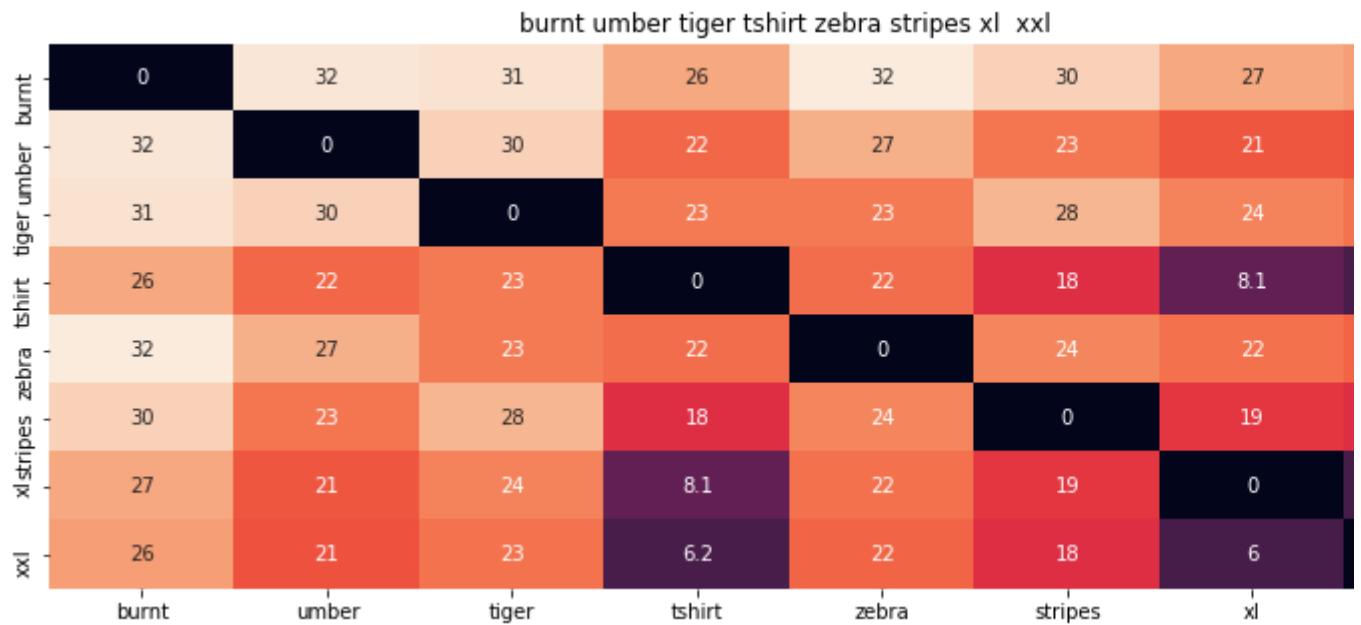
ASIN : B016CU40IY

Brand : Juiceclouds

euclidean distance from input : 6.4117067867524336

When Image feature has maximum wieghtage, Title has 2nd highest weightage
 'Brand Color' has minimum weightage

idf_w2v_brand(12566, 20, 5, 1, 10)

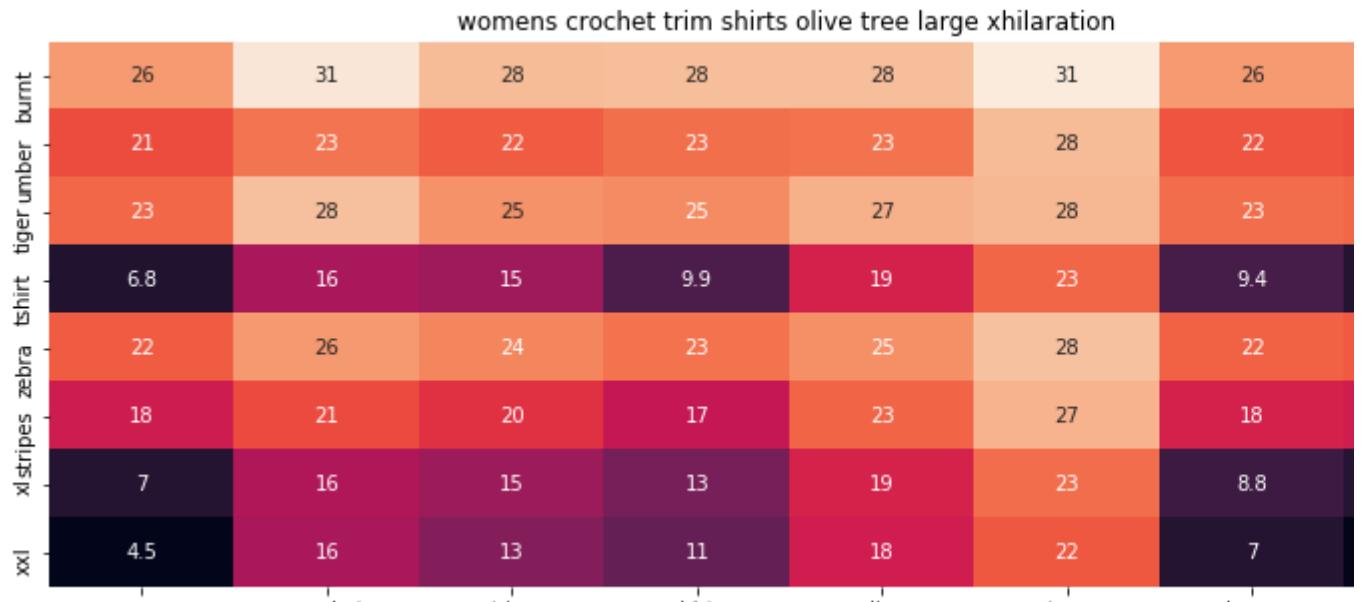


ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.001220703125

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B06XBHN7J

Brand : Xhilaration

euclidean distance from input : 25.897770214125984

	buffalo	david	bitton	nipaw	logo	graphic	tank	white	combo
burnt	31	34	26	26	28	27	26	25	29
tiger	28	28	21	21	24	22	22	20	24
umber	23	32	23	23	25	24	23	22	25
tshirt	21	22	7	7	13	11	9	7.7	14
zebra	21	31	22	22	24	23	22	21	24
stripes	26	28	18	18	19	19	18	17	22
xxl	22	22	7.2	7.2	15	13	9.4	8	13
xxl	20	22	4.7	4.7	14	11	7.8	5.9	12

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

	free	people	free	malibu	thermal	henley	pullover	black	small
burnt	28	27	28	27	33	27	31	25	
umber	24	24	24	22	30	21	25	20	
tiger	26	25	26	26	31	25	26	22	
tshirt	13	13	13	13	23	9.3	14	7.6	
zebra	25	23	25	23	30	22	25	21	
stripes	21	21	21	21	29	18	21	17	

xxl	13	14	13	13	23	11	16	7.9
xxl	12	12	12	12	22	8.8	15	5.9

ASIN : B074MXY984

Brand : We The Free

euclidean distance from input : 27.424663305282593

j brand womens pinstripe shirt xs blue								
burnt	29	26	36	26	26	26	26	26
tiger	27	21	29	22	22	21	21	21
umber	27	23	35	23	23	23	23	23
tshirt	16	6.8	26	6	6	7.7	7.7	7.7
zebra	26	22	33	22	22	23	23	23
stripes	23	18	26	17	17	19	19	19
shirt	16	7	27	8.4	8.4	6.9	6.9	6.9

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

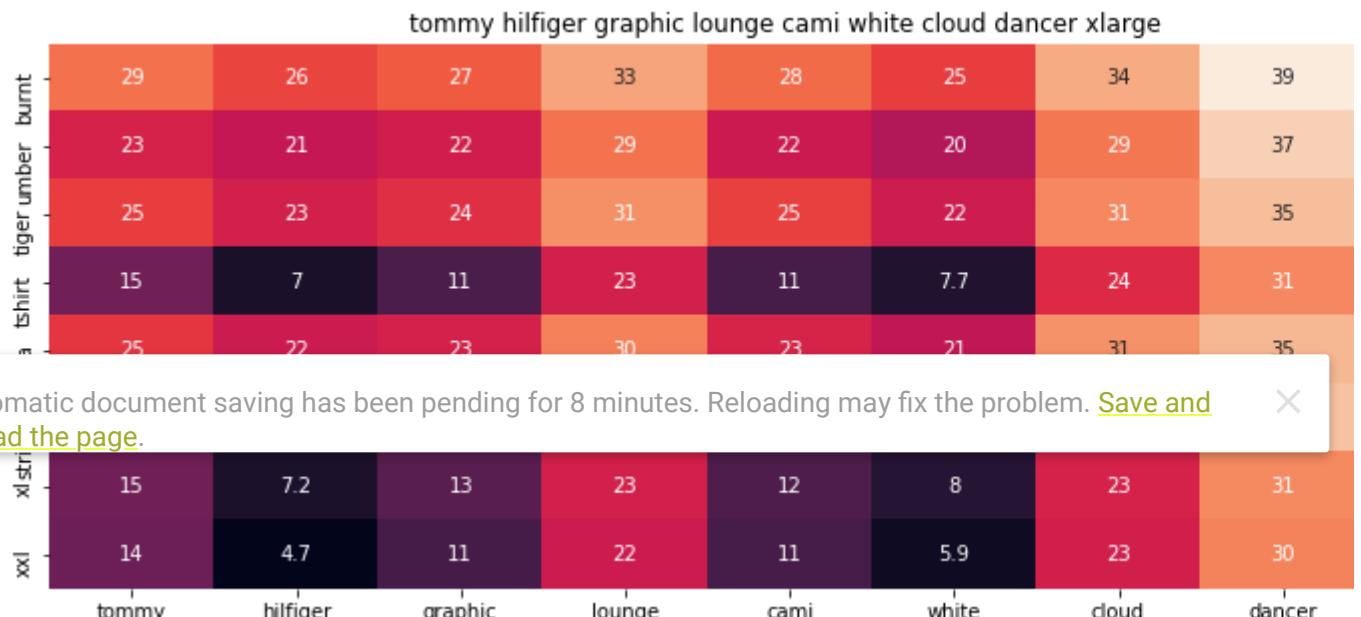
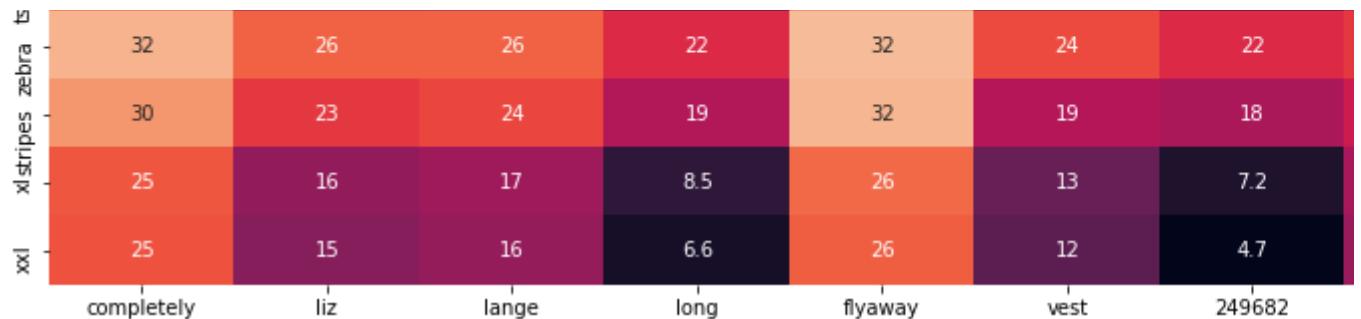
j	brand	womens	pinstripe	shirt
---	-------	--------	-----------	-------

ASIN : B06XYP1X1F

Brand : J Brand Jeans

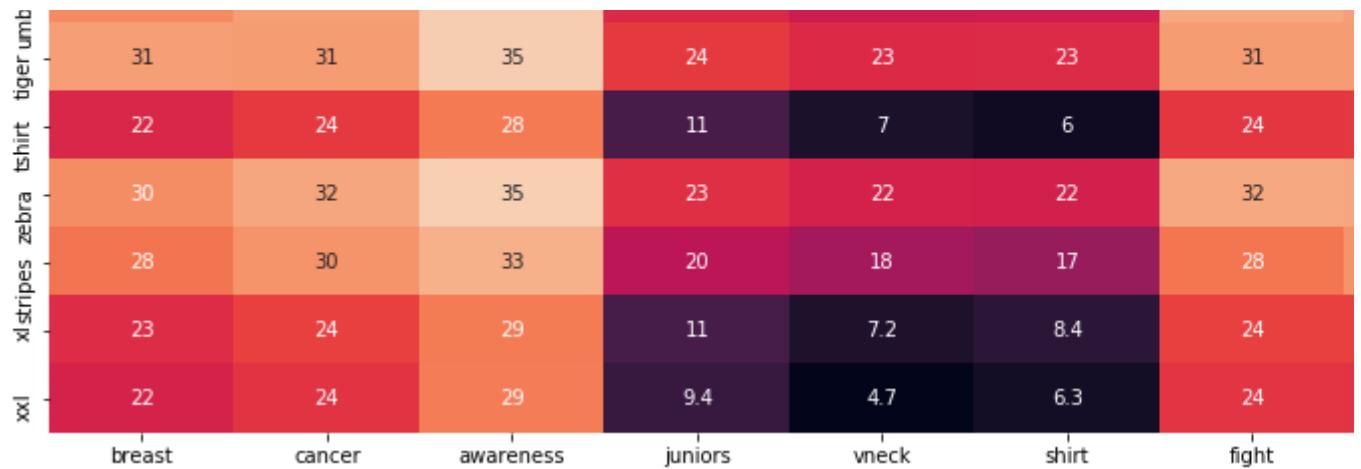
euclidean distance from input : 27.477273968445584

completely liz lange long flyaway vest 249682 turquoise								
burnt	32	30	31	26	36	28	26	26
tiger	33	24	23	22	33	24	21	21
umber	32	27	28	23	34	25	23	23
tshirt	25	15	17	8.6	26	13	7	7



ASIN : B01BMSFYW2
 Brand : igertommy hilf
 euclidean distance from input : 27.827243832337185





ASIN : B016CU40IY

Brand : Juiceclouds

euclidean distance from input : 27.914144952699594



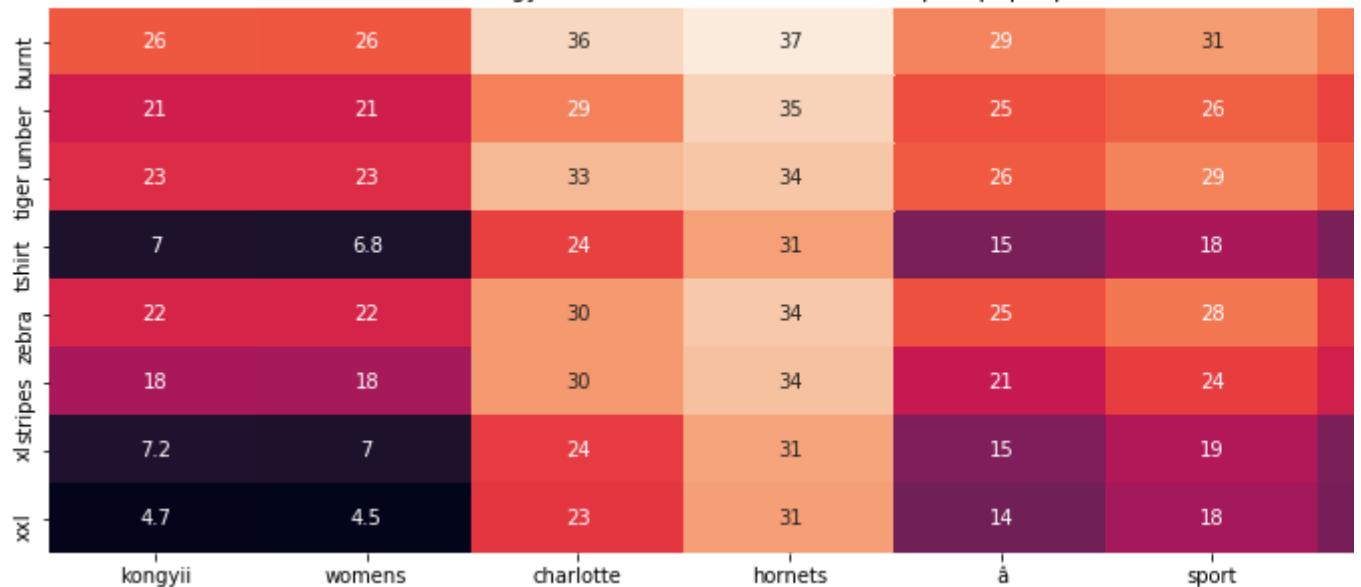
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B01L7ROZNC

Brand : Bila

euclidean distance from input : 28.188810546800546



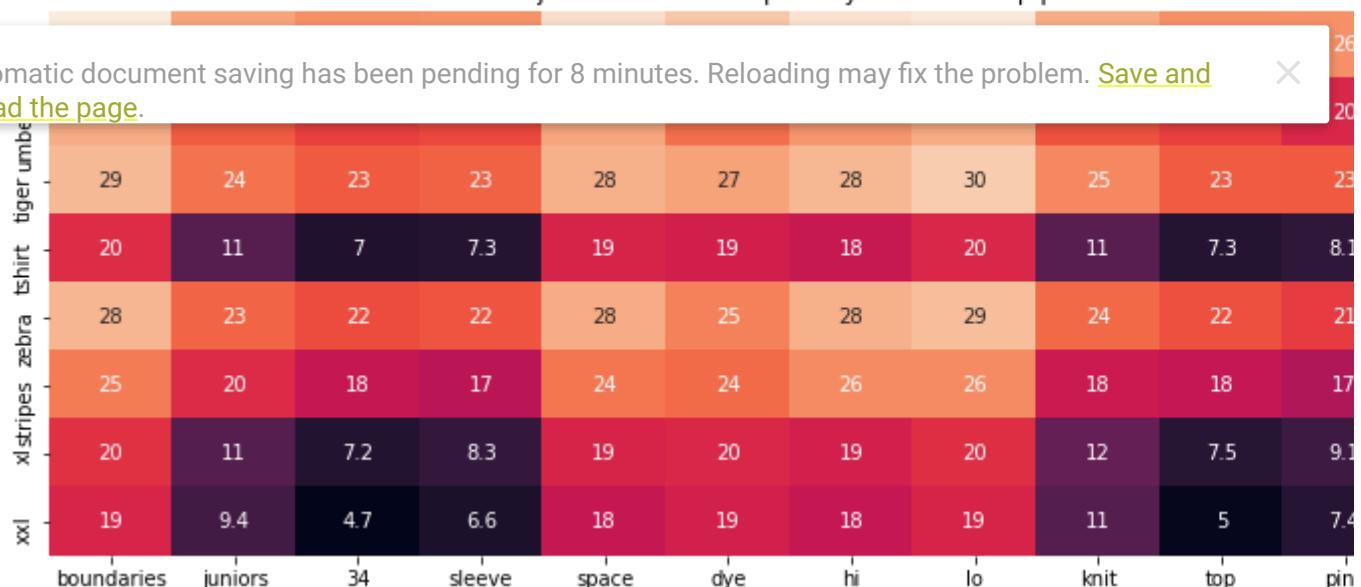
ASIN : B01FJVZST2

Brand : KONGYII

euclidean distance from input : 28.52063341896241

=====

boundaries juniors 34 sleeve space dye hi lo knit top pink mediu



ASIN : B01EXXFS4M

Brand : No Boundaries

euclidean distance from input : 28.787794876143806

=====

salvatore ferragamo geometric print silk top 42 6

	salvatore	ferragamo	geometric	print	silk	top
burnt	26	26	32	26	27	26
tiger	21	21	25	22	21	21
umber	23	23	32	23	24	23
tshirt	7	7	22	9.7	12	7.3
zebra	22	22	27	22	23	22
stripes	18	18	24	19	19	18
xl	7.2	7.2	22	10	13	7.5
xxl	4.7	4.7	21	8.7	11	5

ASIN : B0756JTS1F

Brand : Salvatore Ferragamo

euclidean distance from input : 28.911019802093506

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

girls fairy tail exceed tee shirts black

	girls	fairy	tail	exceed	tee	shirts
burnt	28	33	30	38	27	28
tiger	24	30	26	36	22	23
umber	25	29	26	36	24	25
tshirt	14	25	19	30	11	9.9
zebra	25	29	25	37	23	23
stripes	22	30	22	34	20	17
xl	14	26	19	29	12	13
xxl	13	25	18	29	10	11

ASIN : B01L9F153U

Brand : ATYPEMX

euclidean distance from input : 28.92216058532899

=====

	byoung	womens	henya	womens	light	blue	shirt	size	40l	light
xxl	26	26	26	26	27	26	26	26	26	27
tiger	21	21	21	21	22	20	22	21	21	22
umbre	23	23	23	23	25	23	23	23	23	25
tshirt	7	6.8	7	6.8	13	8	6	8.3	7	13
zebra	22	22	22	22	24	21	22	22	22	24
stripes	18	18	18	18	19	16	17	18	18	19
xl	7.2	7	7.2	7	13	8.7	8.4	7.7	7.2	13
xxl	4.7	4.5	4.7	4.5	12	7	6.3	5.9	4.7	12

ASIN : B06Y41MRCH

Brand : Byoung

euclidean distance from input : 28.924588012740486

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X

	1state	womens	medium	chambray	crochet	solid	blouse
xxl	26	26	26	30	31	27	26
tiger	21	21	22	23	23	23	21
umbre	23	23	24	28	28	25	23
tshirt	7	6.8	9.7	16	16	12	7.3
zebra	22	22	23	25	26	24	22
stripes	18	18	19	21	21	20	18
xl	7.2	7	9.4	17	16	12	8.4
xxl	4.7	4.5	8	16	16	11	6.2

ASIN : B074MK6LV2

Brand : 1.State

euclidean distance from input : 28.97533508102601

=====

stanzino womens long sleeve graphic print plus size top fuchsia xl

	26	26	26	26	27	26	27	26	26	32
burnt	21	21	22	22	22	22	21	21	21	24
tiger	23	23	23	23	24	23	23	23	23	29
tshirt	7	6.8	8.6	7.3	11	9.7	8.4	8.3	7.3	21
zebra	22	22	22	22	23	22	23	22	22	26
stripes	18	18	19	17	19	19	18	18	18	24
xxl	7.2	7	8.5	8.3	13	10	8.1	7.7	7.5	20
	4.7	4.5	6.6	6.6	11	8.7	6.2	5.9	5	20
	stanzino	womens	long	sleeve	graphic	print	plus	size	top	fuchsia

ASIN : B00DP4VHWI

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)



usstore women stripes oversized beach shirt long sleeve casual blouse tee tops free size

	26	26	30	29	28	26	26	26	27	26	27	26
burnt	21	22	23	24	26	22	22	22	22	21	22	22
tiger	23	23	28	26	27	23	23	23	24	23	24	24
tshirt	7	9	18	15	16	6	8.6	7.3	9.9	7.3	11	9.6
zebra	22	23	24	23	25	22	22	22	23	22	23	23
stripes	18	19	0	21	23	17	19	17	19	18	20	18
xxl	7.2	8.7	10	15	17	8.4	8.5	8.7	11	8.4	12	8.3

xxl	7.2	9.3	19	16	17	8.4	8.5	8.3	11	8.4	12	9.8	1
xxl	4.7	7.5	18	15	15	6.3	6.6	6.6	8.9	6.2	10	8.1	1

ASIN : B01DNNI1RO

Brand : Usstore

euclidean distance from input : 29.15367822651612

	hot	sexy	fashion	women	loose	chiffon	short	sleeve	tops	blouse	shirt	
burnt	27	27	27	26	27	29	27	26	26	26	26	
tiger	24	22	22	22	23	21	22	22	22	22	21	
umber	25	24	24	23	24	27	23	23	24	24	23	
tshirt	15	11	10	9	12	15	9.3	7.3	9.6	9.6	7.3	
zebra	25	23	23	23	22	24	23	22	23	23	22	
stripes	21	20	19	19	20	20	19	17	18	18	18	
burnt	15	12	11	9.3	12	16	9.6	8.3	9.8	9.8	8.4	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

ASIN : B00JMAASRO

Brand : Wotefusi

euclidean distance from input : 29.189475019380502

	sexy sheer mesh print long sleeves bodysuit						
burnt	27	28	29	26	26	29	
tiger	22	23	23	22	22	24	
umber	24	25	25	23	23	27	
tshirt	11	13	14	9.7	8.6	14	

	23	25	23	22	22	26	
xxl	20	21	19	19	19	19	
xl	12	13	14	10	8.5	15	
l	10	12	13	8.7	6.6	14	
	sexy	sheer	mesh	print	long	sleeves	

ASIN : B074Z5C98D

Brand : Ariella's closet

euclidean distance from input : 29.238545725117426

	maven	west	striped	sleeveless	lace	peplum	peasant	blouse	yellow	large
xxl	41	29	27	27	27	31	28	26	27	
xl	37	25	20	22	21	22	26	21	20	
l	37	27	23	25	24	28	26	23	24	
m	30	17	11	11	10	17	17	7.3	13	
s	36	25	20	23	23	25	26	22	21	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

xxl	31	17	12	12	11	17	18	8.4	13	
xl	30	16	10	10	9.3	16	17	6.2	12	

ASIN : B01M8GB3AL

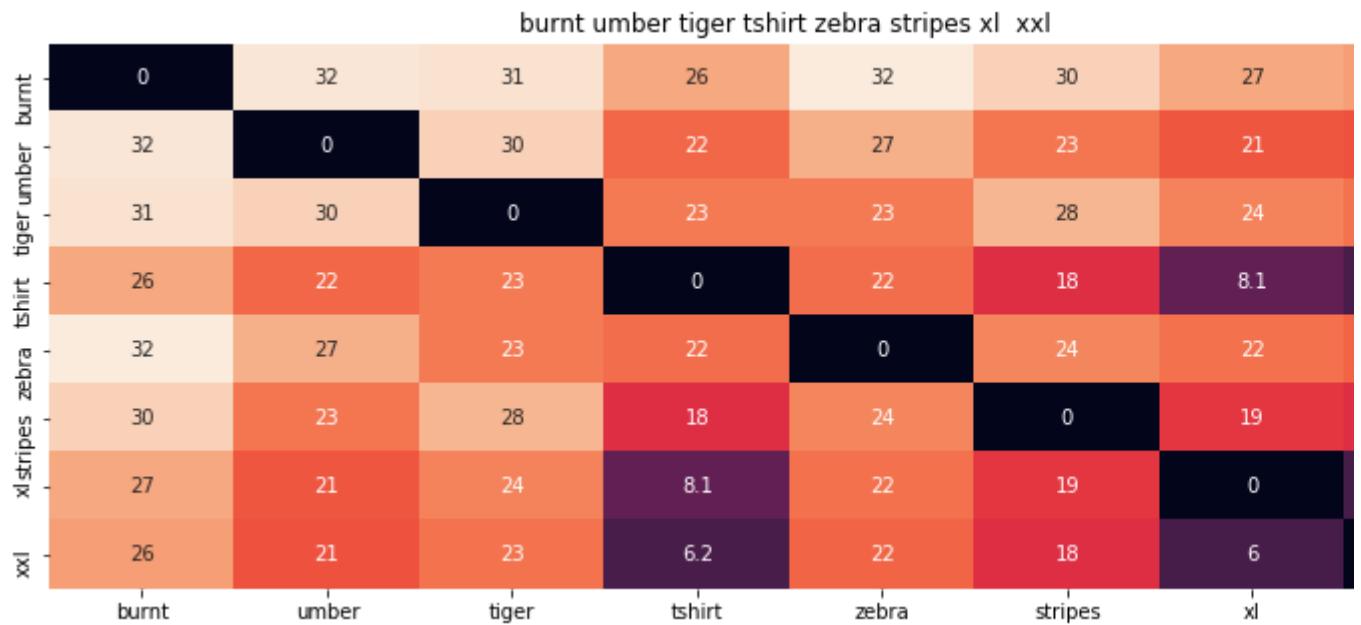
Brand : Maven West

euclidean distance from input : 29.270747232482307

- When Title has max weightage, 'Brand&color' has second highest weightage

```
idf_w2v_brand(12566,20, 10,5,1)
```



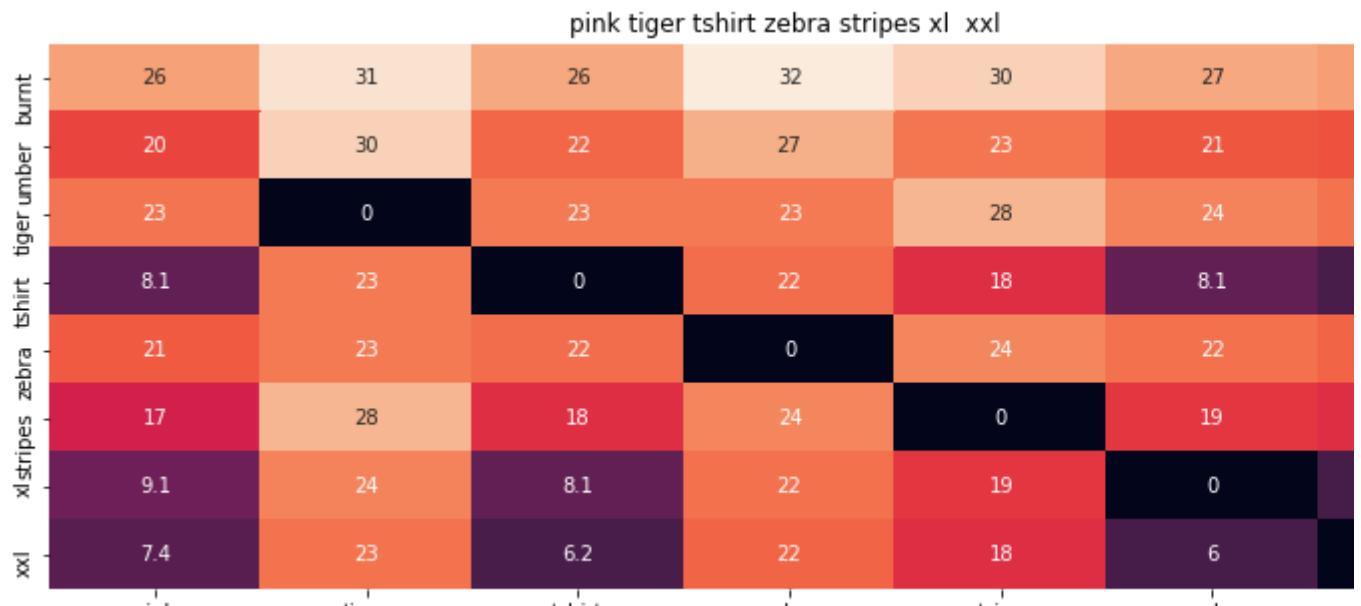


ASIN : B00JXQB5FQ

Brand : Si Row

euclidean distance from input : 0.00244140625

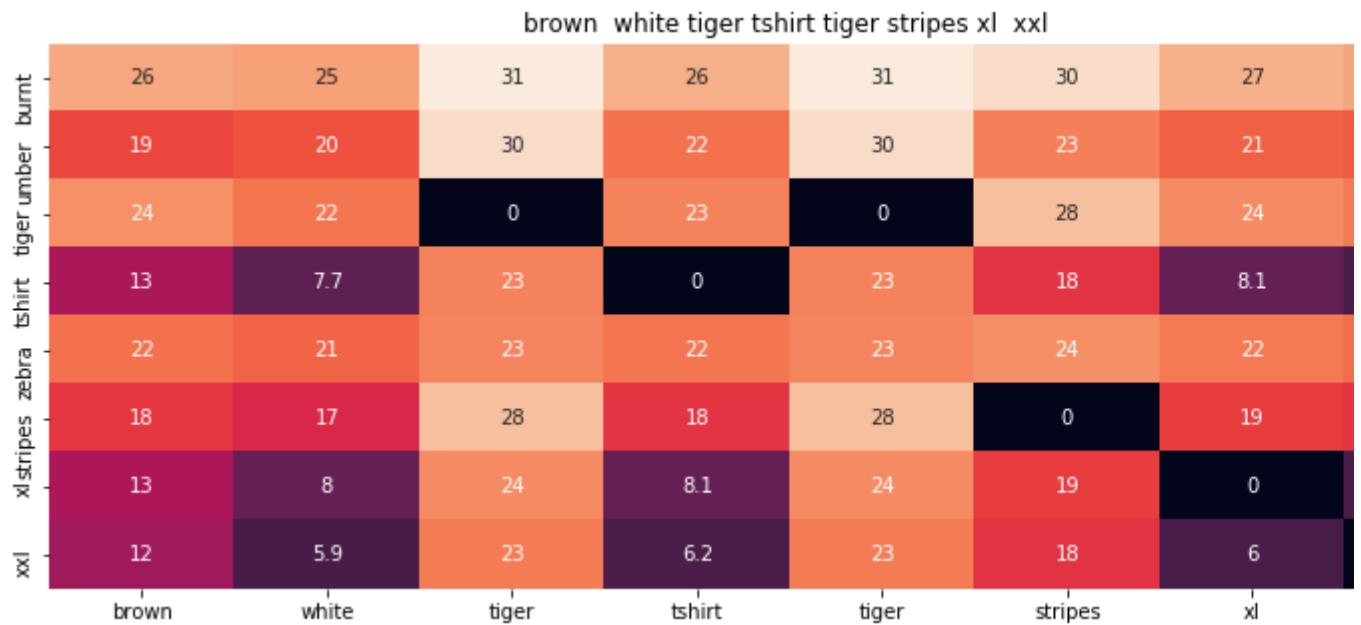
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B00JXQASS6

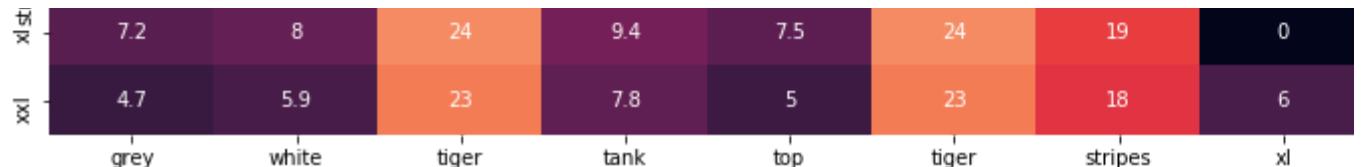
Brand : Si Row

euclidean distance from input : 6.013736963385025



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

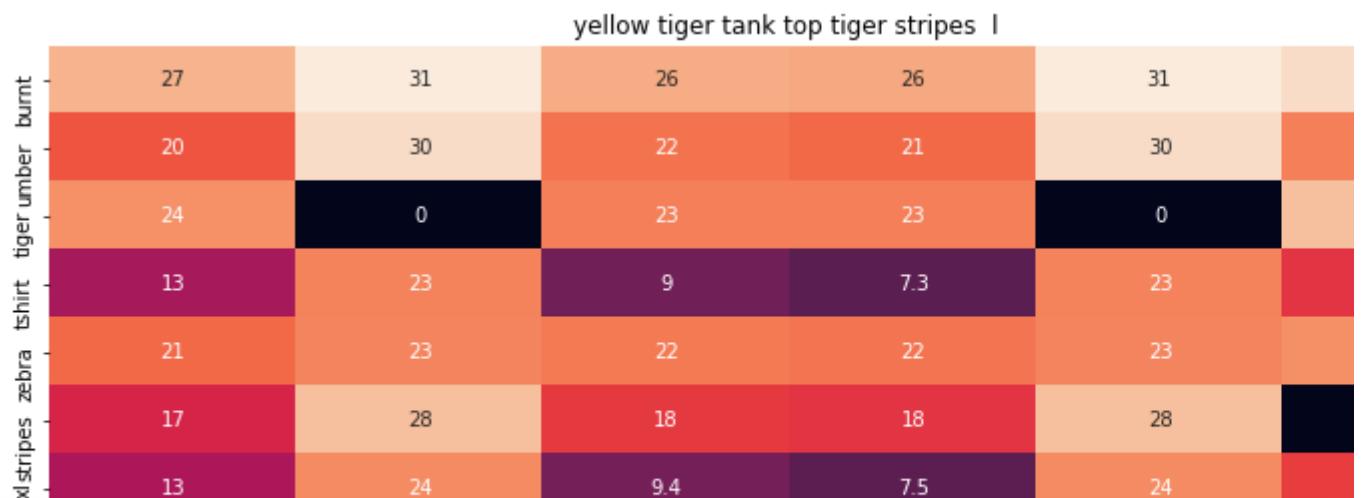




ASIN : B00JXQAFZ2

Brand : Si Row

euclidean distance from input : 7.162451505773941

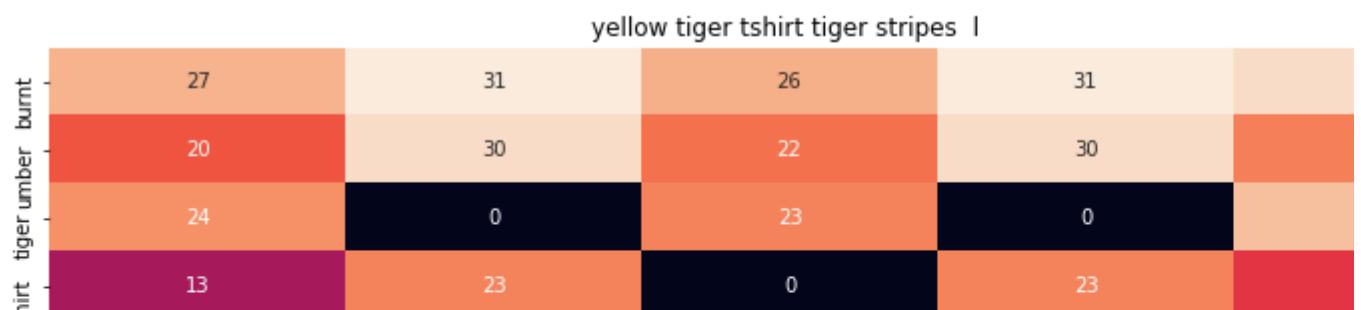


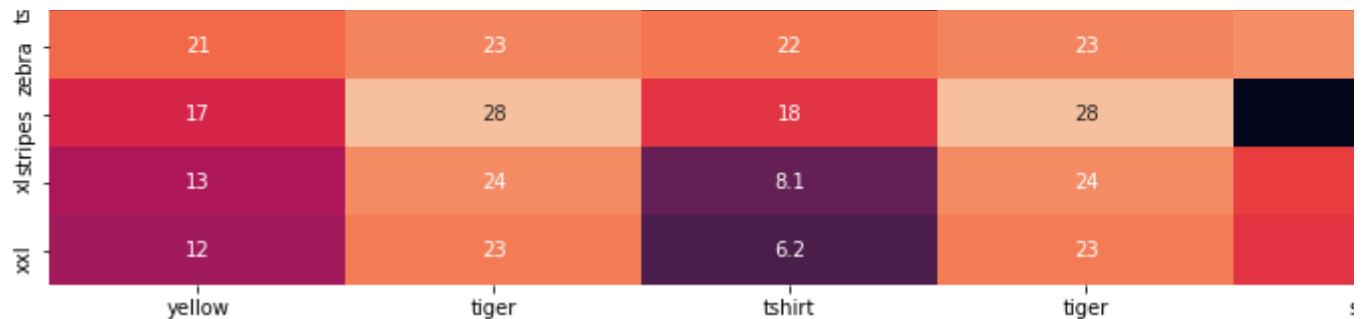
Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

ASIN : B00JXQAUWA

Brand : Si Row

euclidean distance from input : 7.340324640386978

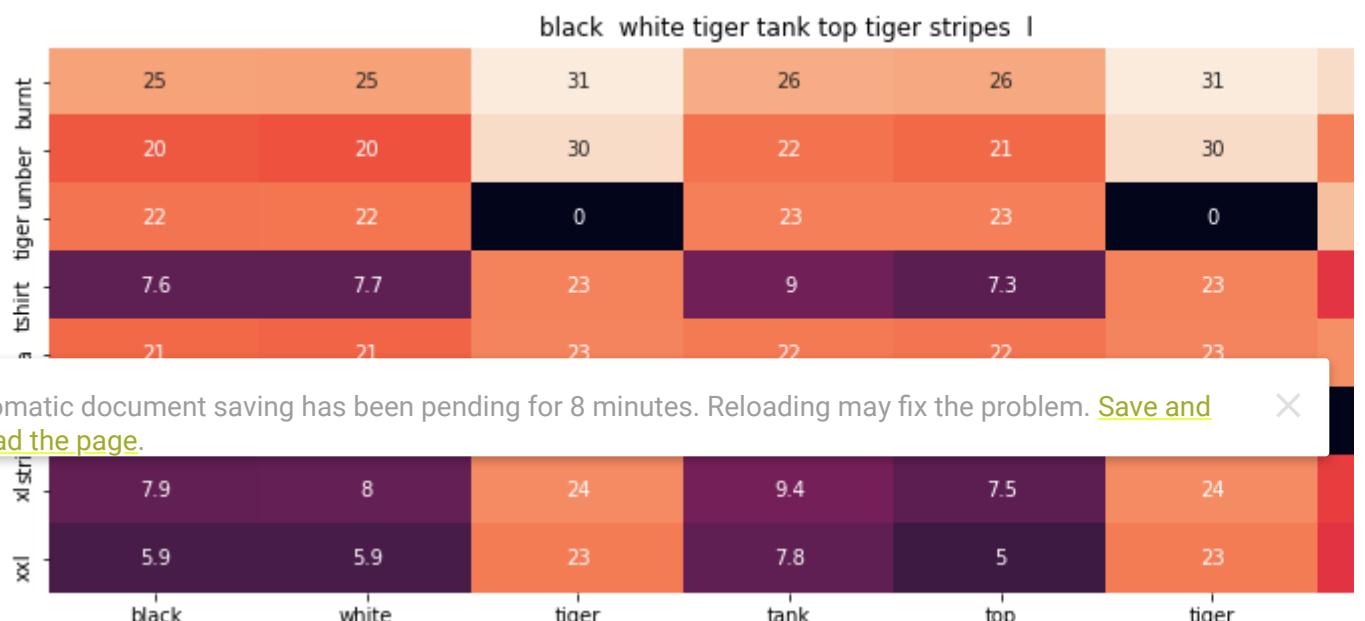




ASIN : B00JXQCUIC

Brand : Si Row

euclidean distance from input : 7.386407137030044



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X



ASIN : B00JXQA094

Brand : Si Row

euclidean distance from input : 7.451913356893936



umb	tiger	tshirt	28	23	24	24	30	23	23	27
umb	tiger	tshirt	21	6.8	13	10	23	10	7	14
umb	tiger	tshirt	27	22	21	21	30	23	22	26
umb	tiger	tshirt	26	18	17	16	27	19	18	19
umb	tiger	tshirt	21	7	13	11	24	11	7.2	15
umb	tiger	tshirt	20	4.5	12	10	23	8.8	4.7	14
	lsu	tigers	colosseum	womens	yellow	purple	slit	back	12	sleeve

ASIN : B073R5Q8HD

Brand : Colosseum

euclidean distance from input : 7.832414664849896

umber	burnt	31	34	26	26	28	27	26	25	29
umber	burnt	28	28	21	21	24	22	22	20	24

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

umber	burnt	31	31	22	22	24	23	22	21	24
umber	burnt	26	28	18	18	19	19	18	17	22
umber	burnt	22	22	7.2	7.2	15	13	9.4	8	13
umber	burnt	20	22	4.7	4.7	14	11	7.8	5.9	12
	burnt	buffalo	david	bitton	nipaw	logo	graphic	tank	white	combo

ASIN : B018H5AZXQ

Brand : Buffalo

euclidean distance from input : 7.994763612973053

burnt	29	25	26	26	27	27	26
tiger	23	20	21	21	20	22	22
umber	26	22	23	23	23	24	23
tshirt	17	7.7	6.8	8.5	11	11	6
zebra	24	21	22	22	20	23	22
stripes	22	17	18	19	16	20	17
xxl	16	8	7	8.4	12	12	8.4
	16	5.9	4.5	6.4	10	10	6.3
	h	bordeaux	white	womens	small	striped	tee

ASIN : B072BVB47Z

Brand : H By Bordeaux

euclidean distance from input : 8.079461144366448

=====

bila size small womens sleeveless blouse red

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

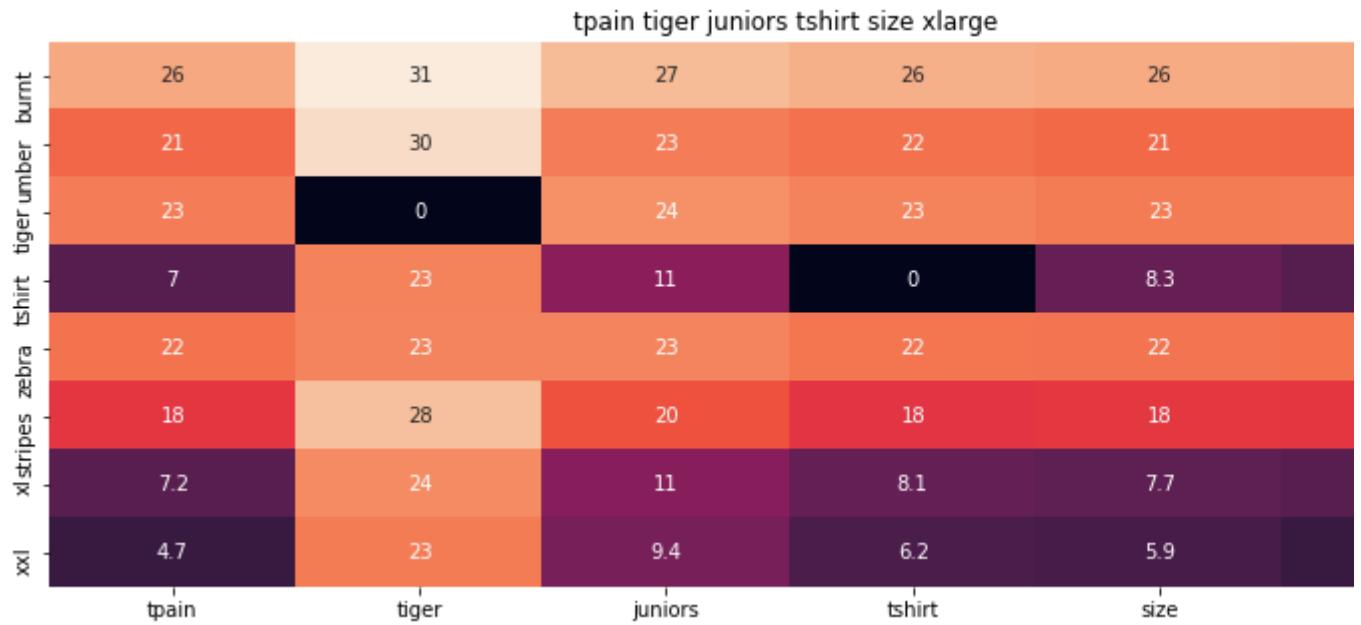
tiger	24	23	23	23	25	23
tshirt	13	8.3	8.5	6.8	11	7.3
zebra	24	22	22	22	23	22
stripes	21	18	19	18	18	18
xxl	12	7.7	8.4	7	12	8.4
	11	5.9	6.4	4.5	10	6.2
	bila	size	small	womens	sleeveless	blouse

ASIN : B01L7ROZNC

Brand : Bila

euclidean distance from input : 8.142929353341717

=====

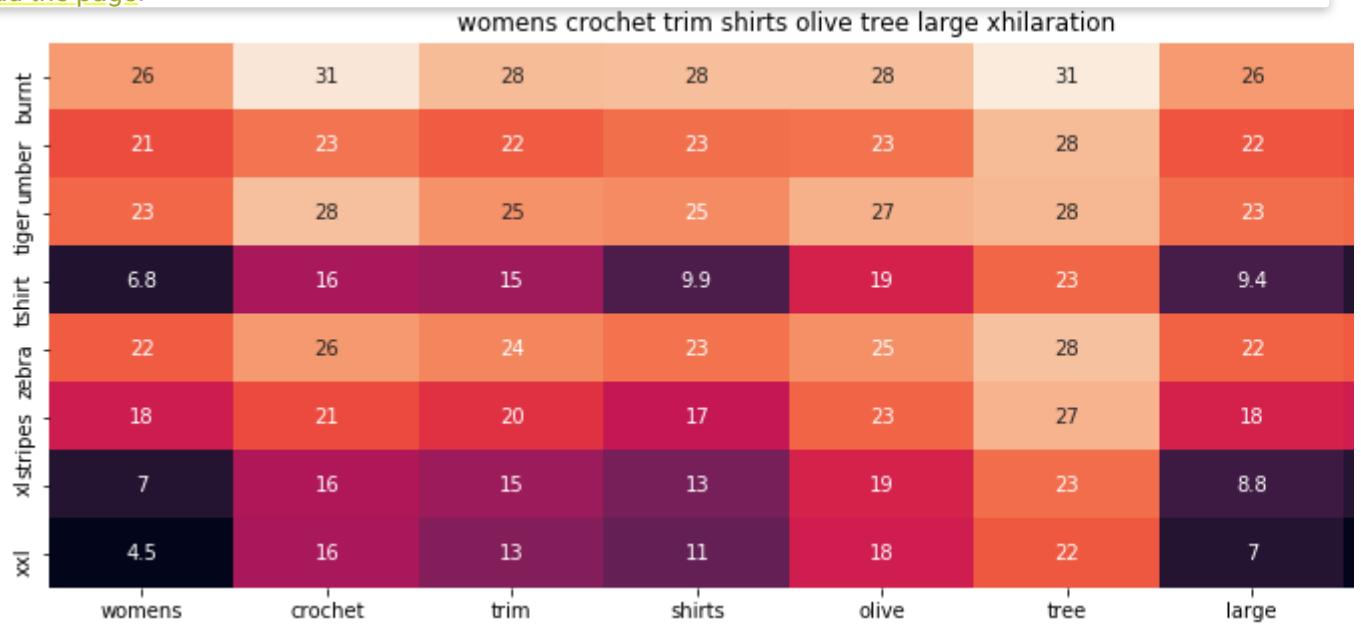


ASIN : B01K0H020G

Brand : Tultex

euclidean distance from input : 8.149006642923016

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#) X



ASIN : B06XBHNM7J

Brand : Xhilaration

euclidean distance from input : 8.165337324368316

kasper teal womens large seamed collar tank blouse blue I								
	kasper	teal	womens	large	seamed	collar	tank	blouse
burnt	26	30	26	26	29	29	26	26
tiger	21	21	21	22	25	23	22	21
umber	23	26	23	23	30	23	23	23
tshirt	7	16	6.8	9.4	19	14	9	7.3
zebra	22	23	22	22	28	22	22	22
xlstripes	18	19	18	18	23	18	18	18
xxl	7.2	17	7	8.8	19	14	9.4	8.4
	4.7	16	4.5	7	18	13	7.8	6.2

ASIN : B0722DJVQP

Brand : Kasper

euclidean distance from input : 8.210880794152875

=====

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#). X

exotic india yellow gray jamawar wrap fauxfur collar								
	exotic	india	yellow	gray	jamawar	wrap	fauxfur	
burnt	29	32	27	27	26	28	26	
tiger	24	26	20	19	21	25	21	
umber	25	29	24	23	23	27	23	
tshirt	17	20	13	12	7	15	7	
zebra	23	29	21	22	22	26	22	
xlstripes	22	27	17	18	18	21	18	
xxl	17	19	13	11	7.2	16	7.2	
	15	19	12	11	4.7	15	4.7	

ASIN : B073ZHRBV8

Brand : Exotic India

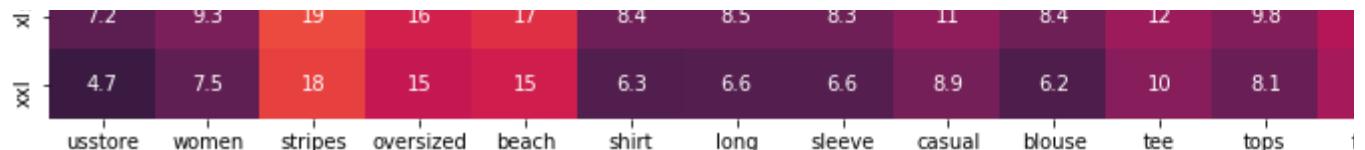
euclidean distance from input : 8.214922189938385

	bobeau	peach	womens	small	tribal	lace	tank	blouse
burnt	26	29	26	26	31	27	26	26
tiger	21	23	21	21	28	21	22	21
umber	23	26	23	23	27	24	23	23
tshirt	7	17	6.8	8.5	21	10	9	7.3
zebra	22	24	22	22	28	23	22	22
stripes	18	22	18	19	26	17	18	18
xxl	7.2	18	7	8.4	22	11	9.4	8.4
xl	4.7	17	4.5	6.4	21	9.3	7.8	6.2

ASIN : B072JTHCX6

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#)

	usstore	women	stripes	oversized	beach	shirt	long	sleeve	casual	blouse	tee	tops	free	size
burnt	26	26	30	29	28	26	26	26	27	26	27	26	26	
tiger	21	22	23	24	26	22	22	22	22	21	22	22	22	
umber	23	23	28	26	27	23	23	23	24	23	24	24	24	
tshirt	7	9	18	15	16	6	8.6	7.3	9.9	7.3	11	9.6		
zebra	22	23	24	23	25	22	22	22	23	22	23	23	23	
stripes	18	19	0	21	23	17	19	17	19	18	20	18		
xxl	7.2	9.7	10	15	17	8.4	8.5	8.7	9.1	8.4	12	8.8	9.2	



ASIN : B01DNNI1RO

Brand : Usstore

euclidean distance from input : 8.26478385947879



Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page](#).

ASIN : B06XTPC3FP

Brand : Kirkland Signature

euclidean distance from input : 8.269013166653473



zebra	22	29	21	22	25	22	23	22	25	21
stripes	18	26	25	21	22	18	20	17	22	17
xl	7.2	19	21	15	19	7.2	13	8.4	16	13
xxl	4.7	18	19	13	18	4.7	11	6.3	15	12
	vansty	im	monkey	cute	tail	100	cotton	shirt	female	yellow

ASIN : B01EFSLHW2

Brand : Vansty

euclidean distance from input : 8.274195708856244

	daniel	rainn	orange	pink	ivory	white	print	chiffon	tank	top	68	white	ivory	xl
burnt	31	26	27	26	29	25	26	29	26	26	26	26	25	
tiger	26	21	20	20	23	20	22	21	22	21	21	21	20	
umber	28	23	25	23	22	22	23	27	23	23	23	23	22	
tshirt	17	7	13	8.1	18	7.7	9.7	15	9	7.3	7	7.7		
bra	26	22	22	21	23	21	22	24	22	22	22	22	21	

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X

xl	17	7.2	13	9.1	18	8	10	16	9.4	7.5	7.2	8	
xxl	17	4.7	13	7.4	17	5.9	8.7	15	7.8	5	4.7	5.9	

ASIN : B01IPV1SFQ

Brand : Daniel Rainn

euclidean distance from input : 8.28563141822815

Conclusion

1. We have considered 16k datapoints for above experiment
2. We have built 'idf-w2v-weightage' , 'Brand & Color' and 'Image feature' weightage feature model.
3. We have given max and min weightage to these features interchanginly.

4. when we gave max importance to 'Title' model recommended items based on Query title.

Automatic document saving has been pending for 8 minutes. Reloading may fix the problem. [Save and reload the page.](#) X