Model and Feature Engineering

In [1]:
```
#import all the necessary packages.
import matplotlib.pyplot as plt
import numpy as np
import pandas as pd
import warnings
warnings.filterwarnings("ignore")
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
import nltk
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 scipy.sparse import hstack
```

In [2]:
```
# loading preprocessed data
data=pd.read_csv(r'preprocessed_with_clusterlabel.csv',index_col=0)
#index_col=0 to get rid of unnamed column
data.head(2)
```

Out[2]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negative	neut
0	garlic oil vegetarian capsule	beauty_hygiene	haircare	srisriayurveda	hairoil_serum	this product contains garlic oil that known he...	220.0	0.0	0.046	0.6
1	water bottle orange	kitchen_garden_pets	storage_accessories	mastercook	water_fridgebottles	each product microwave safe refrigerator safe ...	180.0	0.0	0.000	0.6

In [3]:
```
data.shape
```

Out[3]:
```
(27164, 13)
```

In [4]:
```
data.columns
```

Out[4]:
```
Index(['product', 'category', 'sub_category', 'brand', 'type', 'description',
       'sale_price', 'discount_%', 'negative', 'neutral', 'positive',
       'compound', 'cluster_label'],
      dtype='object')
```

Model-1

`Bag of words` for Preprocessed Product, `One Hot encoding` for categorical features(category,sub_category,brand,type)

In [5]:
```
# Bag of words for product feature
vectorizer=CountVectorizer()
product_bow=vectorizer.fit_transform(data['product'].values)

print("After vectorization")
print('The shape of product feature vector',product_bow.shape)
```

```
After vectorization
The shape of product feature vector (27164, 8687)
```

In [6]:
```
# one hot encoding for category feature
ohe=CountVectorizer()
category_ohe=ohe.fit_transform(data['category'].values)

print("After vectorization")
print('The shape of category ohe vector',category_ohe.shape)
```

```
After vectorization
The shape of category ohe vector (27164, 11)
```

In [7]:
```
# one hot encoding for sub_category feature
ohe=CountVectorizer()
sub_category_ohe=ohe.fit_transform(data['sub_category'].values)

print("After vectorization")
print('The shape of sub_category ohe vector',sub_category_ohe.shape)
```

```
After vectorization
The shape of sub_category ohe vector (27164, 90)
```

In [8]:
```
# one hot encoding for sub_category feature
brandohe=CountVectorizer()
brand_ohe=brandohe.fit_transform(data['brand'].values)

print("After vectorization")
print('The shape of brand ohe vector',brand_ohe.shape)
```

```
After vectorization
The shape of brand ohe vector (27164, 2294)
```

In [9]:
```
# one hot encoding for sub_category feature
ohe=CountVectorizer()
type_ohe=ohe.fit_transform(data['type'].values)

print("After vectorization")
print('The shape of type ohe vector',type_ohe.shape)
```

```
After vectorization
The shape of type ohe vector (27164, 433)
```

In [10]:
```
# concatenating all feature vectors and other numerical value columns(,sale_price,'negative', 'neutral', 'positive','compound', 'cluster_label')
from scipy.sparse import hstack
X_bow = hstack ((product_bow,category_ohe,sub_category_ohe,brand_ohe,type_ohe,data['sale_price'].values.reshape(-1,1), \
               data['negative'].values.reshape(-1,1),data['neutral'].values.reshape(-1,1),data['positive'].values.reshape(-1,1), \
               data['compound'].values.reshape(-1,1),data['cluster_label'].values.reshape(-1,1))).tocsr()
```

In [11]:
```
print('Shape of BOW feature vector:',X_bow.shape)
```

```
Shape of BOW feature vector: (27164, 11521)
```

In [12]:
```
# Model on bag of words for product feature

def bag_of_words_product(prod_index,num_results):

    # prod_index: product index in the given data
    # num_results: number of similar products to show

    # the metric we used here is cosine, the coside distance is mesured as K(X, Y) = <X, Y> / (||X||*||Y||)
    cosine_sim=cosine_similarity(X_bow,X_bow[prod_index])

    # np.argsort will return indices of the nearest products
    indices = np.argsort(cosine_sim.flatten())[-num_results:-1]
    # -1 given to exclude the searched product itself from showing in recommendations as cosinine similarity will be 1 for same product
    # flipping the indices so that the product with more similarity is shown first
    # argsort will do sorting of indices from smallest to largest value
    indices=np.flip(indices) # to get sorting in descending order

    #psimilarity will store the similarity
    psimilarity  = np.sort(cosine_sim.flatten())[-num_results:-1]
    psimilarity = np.flip(psimilarity)

    print('The searched product is:\n',prod_index,":",data['product'].loc[prod_index])
    ##https://appdividend.com/2022/07/27/how-to-print-bold-python-text/
    print('\nTop '+str(num_results-1)+' Similar products for "'+'\033[1m'+data['product'].loc[prod_index]+'\033[0m' +'" are:')
    print("="*70,'\n')

    for i in range (0,len(indices)):
        print(indices[i],":",data['product'].loc[indices[i]])
        print('Cosine similarity:',np.round(psimilarity[i],6))
        print("-"*50,'\n')

    return data.loc[np.append([prod_index],[indices])] # appending prod_index so as we get query product in dataframe

#https://numpy.org/doc/stable/reference/generated/numpy.append.html
```

In [13]:
```
bag_of_words_product(57,11) # without adding discount_%
# ranking of similar recommendations is based on cosine similarity only
```

```
The searched product is:
 57 : argan liquid gold hair spa

Top 10 Similar products for "argan liquid gold hair spa" are:
======================================================================

15854 : cream anti hair loss
Cosine similarity: 0.999901
--------------------------------------------------

18644 : premium henna hair treatment
Cosine similarity: 0.999897
--------------------------------------------------

3738 : apple cider vinegar organic argan oil hair shampoo argan hair conditioner
Cosine similarity: 0.999896
--------------------------------------------------

10754 : argan hair cream
Cosine similarity: 0.999896
--------------------------------------------------

17996 : ultimate hair repair shampoo moroccan argan hair conditioner
Cosine similarity: 0.999894
--------------------------------------------------

20841 : ultra nourishing hair shampoo moroccan argan hair conditioner
Cosine similarity: 0.999894
--------------------------------------------------

23782 : gliss hair repair total repair anti hair breakage treatment
Cosine similarity: 0.999892
--------------------------------------------------

21884 : gliss hair repair ultimate oil elixir structure build treatment
Cosine similarity: 0.99989
--------------------------------------------------

12150 : argan oil conditioner
Cosine similarity: 0.99989
--------------------------------------------------

21762 : gliss hair repair intense therapy bond repair mask
Cosine similarity: 0.999889
--------------------------------------------------
```

Out[13]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negative	neu
57	argan liquid gold hair spa	beauty_hygiene	haircare	aromatreasures	hair_scalptreatment	our beautifully crafted hair spa collection pr...	199.50	5.0	0.054	0.
15854	cream anti hair loss	beauty_hygiene	haircare	himalayawellness	hair_scalptreatment	himalaya anti hair loss cream promotes hair gr...	243.75	25.0	0.068	0.(
18644	premium henna hair treatment	beauty_hygiene	haircare	madilu	hair_scalptreatment	used treat hair fall hair growth early greying...	225.00	0.0	0.000	0.
3738	apple cider vinegar organic argan oil hair sha...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	apple cider vinegar argan shampoo gentle and n...	898.00	0.0	0.004	0.
10754	argan hair cream	beauty_hygiene	haircare	inatur	hair_scalptreatment	inatur moroccan argan oil hair cream contains ...	440.00	20.0	0.000	0.
17996	ultimate hair repair shampoo moroccan argan ha...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	stbotanica moroccan argan oil conditioner affo...	898.00	0.0	0.065	0.(
20841	ultra nourishing hair shampoo moroccan argan h...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	ultra nourishing shampoo proven lock moisture ...	898.00	0.0	0.034	0.
23782	gliss hair repair total repair anti hair break...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	liquid keratin reconstructs the hair and renew...	775.00	0.0	0.102	0.
21884	gliss hair repair ultimate oil elixir structur...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	schwarzkopf gliss hair repair with liquid kera...	775.00	0.0	0.039	0.
12150	argan oil conditioner	beauty_hygiene	haircare	biotique	shampoo_conditioner	biotique argan oil conditioner made from botan...	247.50	25.0	0.039	0.
21762	gliss hair repair intense therapy bond repair ...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	gliss hair repair intense therapy bond repair ...	775.00	0.0	0.038	0.

Taking discount_% for ranking the similar items along with similarity

In [14]:
```
# taking discount_% for ranking the similar items
#https://www.geeksforgeeks.org/sort-rows-or-columns-in-pandas-dataframe-based-on-values/
def bag_of_words_product_with_discount(prod_index,num_results):

    # prod_index: product index in the given data
    # num_results: number of similar products to show

    # the metric we used here is cosine, the coside distance is mesured as K(X, Y) = <X, Y> / (||X||*||Y||)
    cosine_sim=cosine_similarity(X_bow,X_bow[prod_index])

    # np.argsort will return indices of the nearest products
    indices = np.argsort(cosine_sim.flatten())[-num_results:-1]
    # -1 given to exclude the searched product itself from showing in recommendations as cosinine similarity will be 1 for same product
    # flipping the indices so that the product with more similarity is shown first
    # argsort will do sorting of indices from smallest to largest value

    indices=np.flip(indices)
    #psimilarity will store the similarity
    psimilarity  = np.sort(cosine_sim.flatten())[-num_results:-1]
    psimilarity = np.flip(psimilarity)

    print('The searched\Queried product is:\n',prod_index,":",data['product'].loc[prod_index])
    print('\nTop '+str(num_results-1)+' Similar products for "'+'\033[1m'+data['product'].loc[prod_index]+'\033[0m' +'" are:')
    print("="*70,'\n')

    df=data[['product','discount_%']].loc[indices]
    df['discount_%']=df['discount_%']*0.5/100 # multiplied by 0.5 to give half weightage to discount % and divided by 100 to convert
    # percentage to decimal
    df['similarity']=psimilarity.tolist() # adding similarity scores as s new column to df

    df['rank_score']= df['discount_%']+df['similarity'] # creating rank score by adding similarity and discount

    df=df.sort_values(by='rank_score',ascending=False)

    lst=[] # list to store indices after sorting
    for ind in df.index:
        lst.append(ind)
        print(ind,":",df['product'][ind])
        print('Cosine Similarity with queried product is :',np.round(df['similarity'][ind],6))
        print('Discount %: ',np.round(df['discount_%'][ind]/0.5*100,4)) # restoring discount to original scale
        print('-'*50,'\n')
    return data.loc[np.append([prod_index],lst)] # appending prod_index so as we get query product in dataframe
```

In [15]:
```
bow_57=bag_of_words_product_with_discount(57,11)
bow_57
```

```
The searched\Queried product is:
 57 : argan liquid gold hair spa

Top 10 Similar products for "argan liquid gold hair spa" are:
======================================================================

15854 : cream anti hair loss
Cosine Similarity with queried product is : 0.999901
Discount %:  25.0
--------------------------------------------------

12150 : argan oil conditioner
Cosine Similarity with queried product is : 0.99989
Discount %:  25.0
--------------------------------------------------

10754 : argan hair cream
Cosine Similarity with queried product is : 0.999896
Discount %:  20.0
--------------------------------------------------

18644 : premium henna hair treatment
Cosine Similarity with queried product is : 0.999897
Discount %:  0.0
--------------------------------------------------

3738 : apple cider vinegar organic argan oil hair shampoo argan hair conditioner
Cosine Similarity with queried product is : 0.999896
Discount %:  0.0
--------------------------------------------------

17996 : ultimate hair repair shampoo moroccan argan hair conditioner
Cosine Similarity with queried product is : 0.999894
Discount %:  0.0
--------------------------------------------------

20841 : ultra nourishing hair shampoo moroccan argan hair conditioner
Cosine Similarity with queried product is : 0.999894
Discount %:  0.0
--------------------------------------------------

23782 : gliss hair repair total repair anti hair breakage treatment
Cosine Similarity with queried product is : 0.999892
Discount %:  0.0
--------------------------------------------------

21884 : gliss hair repair ultimate oil elixir structure build treatment
Cosine Similarity with queried product is : 0.99989
Discount %:  0.0
--------------------------------------------------

21762 : gliss hair repair intense therapy bond repair mask
Cosine Similarity with queried product is : 0.999889
Discount %:  0.0
--------------------------------------------------
```

Out[15]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negative	neu
57	argan liquid gold hair spa	beauty_hygiene	haircare	aromatreasures	hair_scalptreatment	our beautifully crafted hair spa collection pr...	199.50	5.0	0.054	0.
15854	cream anti hair loss	beauty_hygiene	haircare	himalayawellness	hair_scalptreatment	himalaya anti hair loss cream promotes hair gr...	243.75	25.0	0.068	0.(
12150	argan oil conditioner	beauty_hygiene	haircare	biotique	shampoo_conditioner	biotique argan oil conditioner made from botan...	247.50	25.0	0.039	0.
10754	argan hair cream	beauty_hygiene	haircare	inatur	hair_scalptreatment	inatur moroccan argan oil hair cream contains ...	440.00	20.0	0.000	0.
18644	premium henna hair treatment	beauty_hygiene	haircare	madilu	hair_scalptreatment	used treat hair fall hair growth early greying...	225.00	0.0	0.000	0.
3738	apple cider vinegar organic argan oil hair sha...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	apple cider vinegar argan shampoo gentle and n...	898.00	0.0	0.004	0.
17996	ultimate hair repair shampoo moroccan argan ha...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	stbotanica moroccan argan oil conditioner affo...	898.00	0.0	0.065	0.(
20841	ultra nourishing hair shampoo moroccan argan h...	beauty_hygiene	haircare	stbotanica	shampoo_conditioner	ultra nourishing shampoo proven lock moisture ...	898.00	0.0	0.034	0.
23782	gliss hair repair total repair anti hair break...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	liquid keratin reconstructs the hair and renew...	775.00	0.0	0.102	0.
21884	gliss hair repair ultimate oil elixir structur...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	schwarzkopf gliss hair repair with liquid kera...	775.00	0.0	0.039	0.
21762	gliss hair repair intense therapy bond repair ...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	gliss hair repair intense therapy bond repair ...	775.00	0.0	0.038	0.

In [16]:
```
bag_of_words_product_with_discount(18623,11)
```

```
The searched\Queried product is:
 18623 : cookies italian biscotti

Top 10 Similar products for "cookies italian biscotti" are:
======================================================================

16664 : cookies almond roasted
Cosine Similarity with queried product is : 0.999902
Discount %:  20.0
--------------------------------------------------

3853 : cookies butter pista
Cosine Similarity with queried product is : 0.999924
Discount %:  10.0
--------------------------------------------------

15398 : artisanal cookies seasons greetings
Cosine Similarity with queried product is : 0.999877
Discount %:  10.0
--------------------------------------------------

26398 : peanut butter cookies
Cosine Similarity with queried product is : 0.999891
Discount %:  0.0
--------------------------------------------------

22473 : cookies oats
Cosine Similarity with queried product is : 0.999889
Discount %:  0.0
--------------------------------------------------

10682 : cookies kesar supreme
Cosine Similarity with queried product is : 0.999885
Discount %:  0.0
--------------------------------------------------

4404 : cookies assorted
Cosine Similarity with queried product is : 0.99986
Discount %:  0.0
--------------------------------------------------

15231 : dark chocolate cookies
Cosine Similarity with queried product is : 0.999854
Discount %:  0.0
--------------------------------------------------

15267 : cashew cookies
Cosine Similarity with queried product is : 0.999846
Discount %:  0.0
--------------------------------------------------

17415 : whole wheat cookies choco chip
Cosine Similarity with queried product is : 0.999845
Discount %:  0.0
--------------------------------------------------
```

Out[16]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negativ
18623	cookies italian biscotti	bakery_cakes_dairy	cookies_rusk_khari	lovelybakestudio	bakerybiscuits_cookies	eggless premium quality almonds are twice bake...	166.5	10.0	0.0
16664	cookies almond roasted	bakery_cakes_dairy	cookies_rusk_khari	lovelybakestudio	bakerybiscuits_cookies	eggless premium almonds and goodness butter co...	148.0	20.0	0.0
3853	cookies butter pista	bakery_cakes_dairy	cookies_rusk_khari	lovelybakestudio	bakerybiscuits_cookies	eggless crunchy buttery tasteful baked perfect...	166.5	10.0	0.1
15398	artisanal cookies seasons greetings	bakery_cakes_dairy	cookies_rusk_khari	cookieman	bakerybiscuits_cookies	cookie man presents delightful pack artisanal ...	175.5	10.0	0.0
26398	peanut butter cookies	bakery_cakes_dairy	cookies_rusk_khari	thebakersdozen	bakerybiscuits_cookies	enjoy the baker dozen healthy peanut butter co...	170.0	0.0	0.0
22473	cookies oats	bakery_cakes_dairy	cookies_rusk_khari	bhealthy	bakerybiscuits_cookies	power the with wholesome whole wheat oat cooke...	159.0	0.0	0.0
10682	cookies kesar supreme	bakery_cakes_dairy	cookies_rusk_khari	momkhatai	bakerybiscuits_cookies	momkhatai kesar supreme are handmade cookies w...	160.0	0.0	0.0
4404	cookies assorted	bakery_cakes_dairy	cookies_rusk_khari	momkhatai	bakerybiscuits_cookies	momkhatai assorted cookies contain four delici...	140.0	0.0	0.0
15231	dark chocolate cookies	bakery_cakes_dairy	cookies_rusk_khari	thebakersdozen	premiumcookies	crunchy butter cookies with loads dark chocola...	170.0	0.0	0.0
15267	cashew cookies	bakery_cakes_dairy	cookies_rusk_khari	thebakersdozen	premiumcookies	crunchy butter cookies with cashews chunks the...	150.0	0.0	0.0
17415	whole wheat cookies choco chip	bakery_cakes_dairy	cookies_rusk_khari	bhealthy	bakerybiscuits_cookies	whole wheat cookie filled with choco chips and...	159.0	0.0	0.0

Observatios:

- Even though we searched for cookies its showing similar items milkshake mix
- will see with other featurization techniques

Model-2

Tf-Idf for Product title

In [17]:
```
tfidf_vectorizer = TfidfVectorizer()
tfidf_product = tfidf_vectorizer.fit_transform(data['product'])

print('shape of product title after TF-IDF featurisation:',tfidf_product.shape)
```

```
shape of product title after TF-IDF featurisation: (27164, 8687)
```

In [18]:
```
# concatenating all feature vectors and other numerical value columns(,sale_price,'negative', 'neutral', 'positive','compound', 'cluster_label')

X_tfidf = hstack ((tfidf_product,category_ohe,sub_category_ohe,brand_ohe,type_ohe,data['sale_price'].values.reshape(-1,1), \
               data['negative'].values.reshape(-1,1),data['neutral'].values.reshape(-1,1),data['positive'].values.reshape(-1,1), \
               data['compound'].values.reshape(-1,1),data['cluster_label'].values.reshape(-1,1))).tocsr()
```

In [19]:
```
print('Shape of TF-IDF feature vector:',X_tfidf.shape)
```

```
Shape of TF-IDF feature vector: (27164, 11521)
```

In [20]:
```
def tfidf_product(prod_index,num_results):

    # prod_index: product index in the given data
    # num_results: number of similar products to show

    # the metric we used here is cosine, the coside distance is mesured as K(X, Y) = <X, Y> / (||X||*||Y||)
    cosine_sim=cosine_similarity(X_tfidf,X_tfidf[prod_index])

    # np.argsort will return indices of the nearest products
    indices = np.argsort(cosine_sim.flatten())[-num_results:-1]
    # -1 given to exclude the searched product itself from showing in recommendations as cosinine similarity will be 1 for same product
    # flipping the indices so that the product with more similarity is shown first
    # argsort will do sorting of indices from smallest to largest value

    indices=np.flip(indices)
    #psimilarity will store the similarity
    psimilarity  = np.sort(cosine_sim.flatten())[-num_results:-1]
    psimilarity = np.flip(psimilarity)

    print('The searched\Queried product is:\n',prod_index,":\n",data.loc[prod_index])
    print('\nTop '+str(num_results-1)+' Similar products for "'+'\033[1m'+data['product'].loc[prod_index]+'\033[0m' +'" are:')
    print("="*70,'\n')

    df=data[['product','discount_%']].loc[indices]
    df['discount_%']=df['discount_%']*0.5/100 # multiplied by 0.5 to give half weightage to discount % and divided by 100 to convert
    # percentage to decimal
    df['similarity']=psimilarity.tolist() # adding similarity scores as s new column to df

    df['rank_score']= df['discount_%']+df['similarity'] # creating rank score by adding similarity and discount

    df=df.sort_values(by='rank_score',ascending=False)

    lst=[] # list to store indices after sorting
    for ind in df.index:
        lst.append(ind)
        print(ind,":",df['product'][ind])
        print('Cosine Similarity with queried product is :',np.round(df['similarity'][ind],6))
        print('Discount %: ',np.round(df['discount_%'][ind]/0.5*100,4)) # restoring discount to original scale
        print('-'*50,'\n')

    return data.loc[np.append([prod_index],lst)] # appending prod_index so as we get query product in dataframe
```

In [21]:
```
tfidf_57=tfidf_product(57,11)
tfidf_57
```

```
The searched\Queried product is:
 57 :
 product                                 argan liquid gold hair spa
category                                             beauty_hygiene
sub_category                                               haircare
brand                                                aromatreasures
type                                            hair_scalptreatment
description      our beautifully crafted hair spa collection pr...
sale_price                                                    199.5
discount_%                                                        5
negative                                                      0.054
neutral                                                       0.712
positive                                                      0.233
compound                                                     0.9886
cluster_label                                                     0
Name: 57, dtype: object

Top 10 Similar products for "argan liquid gold hair spa" are:
======================================================================

15854 : cream anti hair loss
Cosine Similarity with queried product is : 0.999959
Discount %:  25.0
--------------------------------------------------

12150 : argan oil conditioner
Cosine Similarity with queried product is : 0.999943
Discount %:  25.0
--------------------------------------------------

26263 : avocado nourish mask for fragile hair
Cosine Similarity with queried product is : 0.999952
Discount %:  20.0
--------------------------------------------------

21421 : argan oil shampoo
Cosine Similarity with queried product is : 0.999942
Discount %:  20.0
--------------------------------------------------

5979 : hibiscus shampoo for dry hair
Cosine Similarity with queried product is : 0.999951
Discount %:  15.0
--------------------------------------------------

12993 : tea tree shampoo for dandruff
Cosine Similarity with queried product is : 0.99995
Discount %:  10.0
--------------------------------------------------

8976 : honey moisture mask for dry damaged hair
Cosine Similarity with queried product is : 0.999958
Discount %:  0.0
--------------------------------------------------

18644 : premium henna hair treatment
Cosine Similarity with queried product is : 0.999957
Discount %:  0.0
--------------------------------------------------

21884 : gliss hair repair ultimate oil elixir structure build treatment
Cosine Similarity with queried product is : 0.999941
Discount %:  0.0
--------------------------------------------------

21762 : gliss hair repair intense therapy bond repair mask
Cosine Similarity with queried product is : 0.999941
Discount %:  0.0
--------------------------------------------------
```

Out[21]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negative	neu
57	argan liquid gold hair spa	beauty_hygiene	haircare	aromatreasures	hair_scalptreatment	our beautifully crafted hair spa collection pr...	199.50	5.0	0.054	0
15854	cream anti hair loss	beauty_hygiene	haircare	himalayawellness	hair_scalptreatment	himalaya anti hair loss cream promotes hair gr...	243.75	25.0	0.068	0
12150	argan oil conditioner	beauty_hygiene	haircare	biotique	shampoo_conditioner	biotique argan oil conditioner made from botan...	247.50	25.0	0.039	0
26263	avocado nourish mask for fragile hair	beauty_hygiene	haircare	godrejprofessional	hair_scalptreatment	with frequent exposure pollution sun and styli...	200.00	20.0	0.028	0
21421	argan oil shampoo	beauty_hygiene	haircare	inatur	shampoo_conditioner	argan oil nutri hydrant shampoo enriched with ...	240.00	20.0	0.023	0
5979	hibiscus shampoo for dry hair	beauty_hygiene	haircare	aromatreasures	shampoo_conditioner	rich formula with natural ingredients which pe...	191.25	15.0	0.000	0
12993	tea tree shampoo for dandruff	beauty_hygiene	haircare	aromatreasures	shampoo_conditioner	thanks its balanced formula which purifies the...	202.50	10.0	0.000	0
8976	honey moisture mask for dry damaged hair	beauty_hygiene	haircare	godrejprofessional	hair_scalptreatment	godrej professional honey moisture mask infuse...	250.00	0.0	0.042	0
18644	premium henna hair treatment	beauty_hygiene	haircare	madilu	hair_scalptreatment	used treat hair fall hair growth early greying...	225.00	0.0	0.000	0
21884	gliss hair repair ultimate oil elixir structur...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	schwarzkopf gliss hair repair with liquid kera...	775.00	0.0	0.039	0
21762	gliss hair repair intense therapy bond repair ...	beauty_hygiene	haircare	schwarzkopf	hair_scalptreatment	gliss hair repair intense therapy bond repair ...	775.00	0.0	0.038	0

Observations:

- We can see that for product index:57 the top 10 similar products using TF-IDF product featurization are different and much better than those of Bag of Words
- Only 3 products are common in both approches

Model-3

TF-IDF weighted Word2Vec for Product title feature

In [22]:
```
# using pre trained word2vec from glove vectors
#https://nlp.stanford.edu/projects/glove/
import pickle
with open('glove_vectors', 'rb') as f:
    model = pickle.load(f)
    glove_words =  set(model.keys())
```

In [23]:
```
# we are converting a dictionary with word as a key, and the idf as a value
dictionary = dict(zip(tfidf_vectorizer.get_feature_names(), list(tfidf_vectorizer.idf_)))
tfidf_words = set(tfidf_vectorizer.get_feature_names())
```

tfidf_W2V vectorization of product feature

In [24]:
```
#vectorizing train  data using tfidf-W2v
# average Word2Vec
# compute tfidf word2vec for each product.
from tqdm import tqdm
product_tfidf_w2v_vectors = []; # the avg-w2v for each product is stored in this list
for product in tqdm(data['product']): # for each product title
    vector = np.zeros(300) # as word vectors are of zero length
    tf_idf_weight =0; # num of words with a valid vector in the product
    for word in product.split(): # for each word in product title
        if (word in glove_words) and (word in tfidf_words):
            vec = model[word] # getting the vector for each word
            # here we are multiplying idf value(dictionary[word]) and the tf value((product.count(word)/len(product.split())))
            tf_idf = dictionary[word]*(product.count(word)/len(product.split())) # getting the tfidf value for each word
            vector += (vec * tf_idf) # calculating tfidf weighted w2v
            tf_idf_weight += tf_idf
    if tf_idf_weight != 0:
        vector /= tf_idf_weight
    product_tfidf_w2v_vectors.append(vector)
```

```
100%|██████████████████████████████| 27164/27164 [00:01<00:00, 15309.58it/s]
```

In [25]:
```
print(len(product_tfidf_w2v_vectors))
print(len(product_tfidf_w2v_vectors[0]))
```

```
27164
300
```

In [26]:
```python
# concatenating all feature vectors and other numerical value columns(,sale_price,'negative', 'neutral', 'positive','compound', 'cluster_label')

X_tfidf_w2v = hstack ((product_tfidf_w2v_vectors,category_ohe,sub_category_ohe,brand_ohe,type_ohe,data['sale_price'].values.reshape(-1,1), \
                       data['negative'].values.reshape(-1,1),data['neutral'].values.reshape(-1,1),data['positive'].values.reshape(-1,1), \
                       data['compound'].values.reshape(-1,1),data['cluster_label'].values.reshape(-1,1))).tocsr()
```

In [27]:
```python
def tfidf_w2v_product(prod_index,num_results):

    # prod_index: product index in the given data
    # num_results: number of similar products to show

    # the metric we used here is cosine, the coside distance is mesured as K(X, Y) = <X, Y> / (||X||*||Y||)
    cosine_sim=cosine_similarity(X_tfidf_w2v,X_tfidf_w2v[prod_index])

    # np.argsort will return indices of the nearest products
    indices = np.argsort(cosine_sim.flatten())[-num_results:-1]
    # -1 given to exclude the searched product itself from showing in recommendations as cosinine similarity will be 1 for same product
    # flipping the indices so that the product with more similarity is shown first
    # argsort will do sorting of indices from smallest to largest value
    indices=np.flip(indices)
    #psimilarity will store the similarity
    psimilarity  = np.sort(cosine_sim.flatten())[-num_results:-1]
    psimilarity = np.flip(psimilarity)

    print('The searched/Queried product is:\n',prod_index,":",data['product'].loc[prod_index])
    print('\nTop '+str(num_results-1)+' Similar products for "'+'\033[1m'+data['product'].loc[prod_index]+'\033[0m' +'" are:')
    print("="*70,'\n')

    df=data[['product','discount_%']].loc[indices]
    df['discount_%']=df['discount_%']*0.5/100 # multiplied by 0.5 to give half weightage to discount % and divided by 100 to convert
    # percentage to decimal
    df['similarity']=psimilarity.tolist() # adding similarity scores as s new column to df

    df['rank_score']= df['discount_%']+df['similarity'] # creating rank score by adding similarity and discount

    df=df.sort_values(by='rank_score',ascending=False)

    lst=[] # list to store indices after sorting
    for ind in df.index:
        lst.append(ind)
        print(ind,":",df['product'][ind])
        print('Cosine Similarity with queried product is :',np.round(df['similarity'][ind],6))
        print('Discount %: ',np.round(df['discount_%'][ind]/0.5*100,4)) # restoring discount to original scale
        print('-'*50,'\n')

    return data.loc[np.append([prod_index],lst)] # appending prod_index so as we get query product in dataframe
```

In [28]:
```python
tfidf_w2v_57=tfidf_w2v_product(57,11)
tfidf_w2v_57
```

```
The searched/Queried product is:
 57 : argan liquid gold hair spa

Top 10 Similar products for "argan liquid gold hair spa" are:
======================================================================

5979 : hibiscus shampoo for dry hair
Cosine Similarity with queried product is : 0.999802
Discount %:  15.0
--------------------------------------------------

18671 : hair oil for strong long thick hair
Cosine Similarity with queried product is : 0.999792
Discount %:  15.0
--------------------------------------------------

17753 : advansed gold coconut hair oil
Cosine Similarity with queried product is : 0.999803
Discount %:  12.0
--------------------------------------------------

16392 : garnier hair color shade natural black loreal paris shampoo
Cosine Similarity with queried product is : 0.999797
Discount %:  10.8434
--------------------------------------------------

13487 : royal gold facial kit for dry skin single time use
Cosine Similarity with queried product is : 0.999814
Discount %:  10.0
--------------------------------------------------

24747 : royal gold facial kit for oily skin single time use
Cosine Similarity with queried product is : 0.99981
Discount %:  10.0
--------------------------------------------------

10565 : royal gold facial kit for oily skin single time use
Cosine Similarity with queried product is : 0.99981
Discount %:  10.0
--------------------------------------------------

21158 : herbal green tea aloevera hair conditioner sls paraben free
Cosine Similarity with queried product is : 0.999816
Discount %:  0.0
--------------------------------------------------

15707 : all natural probiotics shampoo bar for oily hair
Cosine Similarity with queried product is : 0.999809
Discount %:  0.0
--------------------------------------------------

18644 : premium henna hair treatment
Cosine Similarity with queried product is : 0.999802
Discount %:  0.0
--------------------------------------------------
```

Out[28]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	negative	neutra
57	argan liquid gold hair spa	beauty_hygiene	haircare	aromatreasures	hair_scalptreatment	our beautifully crafted hair spa collection pr...	199.50	5.000000	0.054	0.71
5979	hibiscus shampoo for dry hair	beauty_hygiene	haircare	aromatreasures	shampoo_conditioner	rich formula with natural ingredients which pe...	191.25	15.000000	0.000	0.67
18671	hair oil for strong long thick hair	beauty_hygiene	haircare	daburamla	hairoil_serum	dabur amla hair oil infuses your hair with str...	233.75	15.000000	0.000	0.69
17753	advansed gold coconut hair oil	beauty_hygiene	haircare	parachute	hairoil_serum	your hair reflection who you are your identity...	189.20	12.000000	0.043	0.62
16392	garnier hair color shade natural black loreal ...	beauty_hygiene	haircare	bbcombo	haircolor	garnier color naturals shade natural black gar...	222.00	10.843373	0.018	0.69
13487	royal gold facial kit for dry skin single time...	beauty_hygiene	skincare	aromatreasures	facecare	enriched with the goodness gold leaf this trea...	243.00	10.000000	0.031	0.71
24747	royal gold facial kit for oily skin single tim...	beauty_hygiene	skincare	aromatreasures	facecare	enriched with the goodness gold leaf this trea...	243.00	10.000000	0.031	0.71
10565	royal gold facial kit for oily skin single tim...	beauty_hygiene	skincare	aromatreasures	facecare	enriched with the goodness gold leaf this trea...	243.00	10.000000	0.031	0.71
21158	herbal green tea aloevera hair conditioner sls...	beauty_hygiene	haircare	khadinatural	shampoo_conditioner	this green tea aloe vera conditioner nourishes...	250.00	0.000000	0.101	0.59
15707	all natural probiotics shampoo bar for oily hair	beauty_hygiene	haircare	golisoda	shampoo_conditioner	the shampoo bar contains brahmi powder which t...	250.00	0.000000	0.037	0.73
18644	premium henna hair treatment	beauty_hygiene	haircare	madilu	hair_scalptreatment	used treat hair fall hair growth early greying...	225.00	0.000000	0.000	0.72

In [29]:
```python
tfidf_w2v_product(18623,11)
```

```
The searched/Queried product is:
 18623 : cookies italian biscotti

Top 10 Similar products for "cookies italian biscotti" are:
======================================================================

3031 : organic foxtail millet italian thinai rice
Cosine Similarity with queried product is : 0.99952
Discount %:  30.5019
--------------------------------------------------

6982 : organic foxtail millet italian thinai rice
Cosine Similarity with queried product is : 0.99952
Discount %:  30.5019
--------------------------------------------------

14094 : fresho signature plain bread stick happy chef pasta sauce arrabiata
Cosine Similarity with queried product is : 0.999608
Discount %:  17.0561
--------------------------------------------------

13810 : fresho signature plain bread stick happy chef pasta sauce arrabiata
Cosine Similarity with queried product is : 0.999574
Discount %:  17.0561
--------------------------------------------------

6198 : assorted cookies fruit nut choc chip
Cosine Similarity with queried product is : 0.999537
Discount %:  15.0
--------------------------------------------------

3853 : cookies butter pista
Cosine Similarity with queried product is : 0.999553
Discount %:  10.0
--------------------------------------------------

17415 : whole wheat cookies choco chip
Cosine Similarity with queried product is : 0.999566
Discount %:  0.0
--------------------------------------------------

15231 : dark chocolate cookies
Cosine Similarity with queried product is : 0.999547
Discount %:  0.0
--------------------------------------------------

15845 : artisanal cookies
Cosine Similarity with queried product is : 0.999525
Discount %:  0.0
--------------------------------------------------

4322 : quadratini wafer cookies chocolate
Cosine Similarity with queried product is : 0.999508
Discount %:  0.0
--------------------------------------------------
```

Out[29]:

	product	category	sub_category	brand	type	description	sale_price	discount_%	nega
18623	cookies italian biscotti	bakery_cakes_dairy	cookies_rusk_khari	lovelybakestudio	bakerybiscuits_cookies	eggless premium quality almonds are twice bake...	166.50	10.000000	0
3031	organic foxtail millet italian thinai rice	foodgrains_oil_masala	organicstaples	bbroyal	organicmillet_flours	foxtail italian millet one those forgotten gra...	180.00	30.501931	0
6982	organic foxtail millet italian thinai rice	foodgrains_oil_masala	dals_pulses	bbroyal	organicmillet_flours	foxtail italian millet one those forgotten gra...	180.00	30.501931	0
14094	fresho signature plain bread stick happy chef ...	bakery_cakes_dairy	cookies_rusk_khari	bbcombo	bakerybiscuits_cookies	cherish the heavenly taste these incredibly de...	177.50	17.056075	0
13810	fresho signature plain bread stick happy chef ...	bakery_cakes_dairy	bakerysnacks	bbcombo	bakerybiscuits_cookies	cherish the heavenly taste these incredibly de...	177.50	17.056075	0
6198	assorted cookies fruit nut choc chip	bakery_cakes_dairy	cookies_rusk_khari	cookieman	bakerybiscuits_cookies	cookie man presents delightful pack assorted a...	165.75	15.000000	0
3853	cookies butter pista	bakery_cakes_dairy	cookies_rusk_khari	lovelybakestudio	bakerybiscuits_cookies	eggless crunchy buttery tasteful baked perfect...	166.50	10.000000	0
17415	whole wheat cookies choco chip	bakery_cakes_dairy	cookies_rusk_khari	bhealthy	bakerybiscuits_cookies	whole wheat cookie filled with choco chips and...	159.00	0.000000	0
15231	dark chocolate cookies	bakery_cakes_dairy	cookies_rusk_khari	thebakersdozen	premiumcookies	crunchy butter cookies with loads dark chocola...	170.00	0.000000	0
15845	artisanal cookies	bakery_cakes_dairy	cookies_rusk_khari	cookieman	bakerybiscuits_cookies	cookie man presents delightful pack artisanal ...	370.00	0.000000	0
4322	quadratini wafer cookies chocolate	gourmet_worldfood	chocolates_biscuits	loacker	cookies_biscotti_wafer	loacker chocolates are made using unique recip...	349.00	0.000000	0

Observations:

- TF-IDF Word2Vec is recommending similar items which are more relavant to search query than BOW
- Overall TF-IDF Word2Vec will be better solution

Summary of all three featurizations techniques for Product title

In [34]:
```python
#Summary of all three featurizations techniques for Product title for product index 57
#https://pandas.pydata.org/docs/user_guide/style.html
#https://stackoverflow.com/questions/59535426/can-you-change-the-caption-font-size-using-pandas-styling
#https://datascientyst.com/set-caption-customize-font-size-color-in-pandas-dataframe/
prod_index=57
d=pd.DataFrame({'BOW':bow_57['product'].values[1:],'TF-IDF':tfidf_57['product'].values[1:],'TF-IDF_W2V':tfidf_w2v_57['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                        'props':[('color','black'),
                                                                                                                 ('font-weight', 'bold'),
                                                                                                                 ('text-align','left'),
                                                                                                                 ('font-size', '16px')]}])
```

Out[34]:

Similar products for : " argan liquid gold hair spa"

	BOW	TF-IDF	TF-IDF_W2V
0	cream anti hair loss	cream anti hair loss	hibiscus shampoo for dry hair
1	argan oil conditioner	argan oil conditioner	hair oil for strong long thick hair
2	argan hair cream	avocado nourish mask for fragile hair	advansed gold coconut hair oil
3	premium henna hair treatment	argan oil shampoo	garnier hair color shade natural black loreal paris shampoo
4	apple cider vinegar organic argan oil hair shampoo argan hair conditioner	hibiscus shampoo for dry hair	royal gold facial kit for dry skin single time use
5	ultimate hair repair shampoo moroccan argan hair conditioner	tea tree shampoo for dandruff	royal gold facial kit for oily skin single time use
6	ultra nourishing hair shampoo moroccan argan hair conditioner	honey moisture mask for dry damaged hair	royal gold facial kit for oily skin single time use
7	gliss hair repair total repair anti hair breakage treatment	premium henna hair treatment	herbal green tea aloevera hair conditioner sls paraben free
8	gliss hair repair ultimate oil elixir structure build treatment	gliss hair repair ultimate oil elixir structure build treatment	all natural probiotics shampoo bar for oily hair
9	gliss hair repair intense therapy bond repair mask	gliss hair repair intense therapy bond repair mask	premium henna hair treatment

In [31]:
```python
#https://pypi.org/project/prettytable/
from prettytable import PrettyTable

table=PrettyTable()

# adding title to table
table.title='Similar Products for: '+ tfidf_57['product'].values[0]
#adding columns to pretty table
table.add_column("TF-IDF",tfidf_57['product'].values[1:]) # indexing from [1:] first to end as 0th index has searched product
table.add_column("TF-IDF_W2V",tfidf_w2v_57['product'].values[1:])

table.align="l"
table._max_width = {"TF-IDF" :61, "TF-IDF_W2V" :64} # define width of columns
#https://stackoverflow.com/questions/59823203/how-can-i-define-the-width-of-a-column-prettytable-python

print(table)
```

```
+-----------------------------------------------------------------------------------------------------------+
|                                  Similar Products for: argan liquid gold hair spa                          |
+-------------------------------------------------------------+---------------------------------------------+
| TF-IDF                                                      | TF-IDF_W2V                                  |
+-------------------------------------------------------------+---------------------------------------------+
| cream anti hair loss                                        | hibiscus shampoo for dry hair               |
| argan oil conditioner                                       | hair oil for strong long thick hair         |
| avocado nourish mask for fragile hair                       | advansed gold coconut hair oil              |
| argan oil shampoo                                           | garnier hair color shade natural black loreal paris shampoo |
| hibiscus shampoo for dry hair                               | royal gold facial kit for dry skin single time use |
| tea tree shampoo for dandruff                               | royal gold facial kit for oily skin single time use |
| honey moisture mask for dry damaged hair                    | royal gold facial kit for oily skin single time use |
| premium henna hair treatment                                | herbal green tea aloevera hair conditioner sls paraben free |
| gliss hair repair ultimate oil elixir structure build       | all natural probiotics shampoo bar for oily hair |
| treatment                                                   |                                             |
| gliss hair repair intense therapy bond repair mask          | premium henna hair treatment                |
+-------------------------------------------------------------+---------------------------------------------+
```

- **Randomly checking for other products**

In [35]:
```python
# checking for prod_index 18623 , top 10 similar items
bow_18623=bag_of_words_product_with_discount(18623,11)
tfidf_18623=tfidf_product(18623,11)
tfidf_w2v_18623=tfidf_w2v_product(18623,11)
```

```
The searched\Queried product is:
 18623 : cookies italian biscotti

Top 10 Similar products for "cookies italian biscotti" are:
======================================================================

16664 : cookies almond roasted
Cosine Similarity with queried product is : 0.999902
Discount %:  20.0
--------------------------------------------------

3853 : cookies butter pista
Cosine Similarity with queried product is : 0.999924
Discount %:  10.0
--------------------------------------------------

15398 : artisanal cookies seasons greetings
Cosine Similarity with queried product is : 0.999877
Discount %:  10.0
--------------------------------------------------

26398 : peanut butter cookies
Cosine Similarity with queried product is : 0.999891
Discount %:  0.0
--------------------------------------------------

22473 : cookies oats
Cosine Similarity with queried product is : 0.999889
Discount %:  0.0
--------------------------------------------------

10682 : cookies kesar supreme
Cosine Similarity with queried product is : 0.999885
Discount %:  0.0
--------------------------------------------------

4404 : cookies assorted
Cosine Similarity with queried product is : 0.99986
Discount %:  0.0
--------------------------------------------------

15231 : dark chocolate cookies
Cosine Similarity with queried product is : 0.999854
Discount %:  0.0
--------------------------------------------------

15267 : cashew cookies
Cosine Similarity with queried product is : 0.999846
Discount %:  0.0
--------------------------------------------------

17415 : whole wheat cookies choco chip
Cosine Similarity with queried product is : 0.999845
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 18623 :
 product                                      cookies italian biscotti
category                                          bakery_cakes_dairy
sub_category                                      cookies_rusk_khari
brand                                               lovelybakestudio
type                                          bakerybiscuits_cookies
description       eggless premium quality almonds are twice bake...
sale_price                                                     166.5
discount_%                                                        10
negative                                                           0
neutral                                                        0.845
positive                                                       0.155
compound                                                      0.8834
cluster_label                                                      7
Name: 18623, dtype: object

Top 10 Similar products for "cookies italian biscotti" are:
======================================================================

21724 : cookies honey corn flakes
Cosine Similarity with queried product is : 0.999912
Discount %:  40.1606
--------------------------------------------------

16664 : cookies almond roasted
Cosine Similarity with queried product is : 0.999954
Discount %:  20.0
--------------------------------------------------

14094 : fresho signature plain bread stick happy chef pasta sauce arrabiata
Cosine Similarity with queried product is : 0.999925
Discount %:  17.0561
--------------------------------------------------

6198 : assorted cookies fruit nut choc chip
Cosine Similarity with queried product is : 0.999932
Discount %:  15.0
--------------------------------------------------

3853 : cookies butter pista
Cosine Similarity with queried product is : 0.999968
Discount %:  10.0
--------------------------------------------------

15398 : artisanal cookies seasons greetings
Cosine Similarity with queried product is : 0.999933
Discount %:  10.0
--------------------------------------------------

26398 : peanut butter cookies
Cosine Similarity with queried product is : 0.999936
Discount %:  0.0
--------------------------------------------------

10682 : cookies kesar supreme
Cosine Similarity with queried product is : 0.99993
Discount %:  0.0
--------------------------------------------------

17415 : whole wheat cookies choco chip
Cosine Similarity with queried product is : 0.999929
Discount %:  0.0
--------------------------------------------------

22473 : cookies oats
Cosine Similarity with queried product is : 0.999919
Discount %:  0.0
--------------------------------------------------

The searched/Queried product is:
 18623 : cookies italian biscotti

Top 10 Similar products for "cookies italian biscotti" are:
======================================================================

3031 : organic foxtail millet italian thinai rice
Cosine Similarity with queried product is : 0.99952
Discount %:  30.5019
--------------------------------------------------

6982 : organic foxtail millet italian thinai rice
Cosine Similarity with queried product is : 0.99952
Discount %:  30.5019
--------------------------------------------------

14094 : fresho signature plain bread stick happy chef pasta sauce arrabiata
Cosine Similarity with queried product is : 0.999608
Discount %:  17.0561
--------------------------------------------------

13810 : fresho signature plain bread stick happy chef pasta sauce arrabiata
Cosine Similarity with queried product is : 0.999574
Discount %:  17.0561
--------------------------------------------------

6198 : assorted cookies fruit nut choc chip
Cosine Similarity with queried product is : 0.999537
Discount %:  15.0
--------------------------------------------------

3853 : cookies butter pista
Cosine Similarity with queried product is : 0.999553
Discount %:  10.0
--------------------------------------------------

17415 : whole wheat cookies choco chip
Cosine Similarity with queried product is : 0.999566
Discount %:  0.0
--------------------------------------------------

15231 : dark chocolate cookies
Cosine Similarity with queried product is : 0.999547
Discount %:  0.0
--------------------------------------------------

15845 : artisanal cookies
Cosine Similarity with queried product is : 0.999525
Discount %:  0.0
--------------------------------------------------

4322 : quadratini wafer cookies chocolate
Cosine Similarity with queried product is : 0.999508
Discount %:  0.0
--------------------------------------------------
```

In [36]:
```python
#Summarizing the results of all 3 methods
prod_index=18623
d=pd.DataFrame({'BOW':bow_18623['product'].values[1:],'TF-IDF':tfidf_18623['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_18623['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                        'props':[('color','black'),
                                                                                                                 ('font-weight', 'bold'),
                                                                                                                 ('text-align','left'),
                                                                                                                 ('font-size', '16px')]}])
```

Out[36]:

Similar products for : " cookies italian biscotti"

	BOW	TF-IDF	TF-IDF_W2V
0	cookies almond roasted	cookies honey corn flakes	organic foxtail millet italian thinai rice
1	cookies butter pista	cookies almond roasted	organic foxtail millet italian thinai rice
2	artisanal cookies seasons greetings	fresho signature plain bread stick happy chef pasta sauce arrabiata	fresho signature plain bread stick happy chef pasta sauce arrabiata
3	peanut butter cookies	assorted cookies fruit nut choc chip	fresho signature plain bread stick happy chef pasta sauce arrabiata
4	cookies oats	cookies butter pista	assorted cookies fruit nut choc chip
5	cookies kesar supreme	artisanal cookies seasons greetings	cookies butter pista
6	cookies assorted	peanut butter cookies	whole wheat cookies choco chip
7	dark chocolate cookies	cookies kesar supreme	dark chocolate cookies
8	cashew cookies	whole wheat cookies choco chip	artisanal cookies
9	whole wheat cookies choco chip	cookies oats	quadratini wafer cookies chocolate

In [39]:
```python
# checking for prod_index 22349 , top 10 similar items
bow_22349=bag_of_words_product_with_discount(22349,11)
tfidf_22349=tfidf_product(22349,11)
tfidf_w2v_22349=tfidf_w2v_product(22349,11)
```

```
The searched\Queried product is:
 22349 : keratin smooth conditioner

Top 10 Similar products for "keratin smooth conditioner" are:
======================================================================

14101 : nourish replenish conditioner
Cosine Similarity with queried product is : 0.999926
Discount %:  25.0
--------------------------------------------------

20866 : climate protection conditioner
Cosine Similarity with queried product is : 0.9999
Discount %:  25.0
--------------------------------------------------

64 : aqua halo rejuvenating conditioner
Cosine Similarity with queried product is : 0.999873
Discount %:  25.0
--------------------------------------------------

26061 : hair fall defense conditioner
Cosine Similarity with queried product is : 0.99987
Discount %:  22.8663
--------------------------------------------------

924 : oil nourish conditioner
Cosine Similarity with queried product is : 0.999884
Discount %:  20.1005
--------------------------------------------------

15459 : color protect conditioner
Cosine Similarity with queried product is : 0.999884
Discount %:  20.0
--------------------------------------------------

11821 : aloe hair conditioner
Cosine Similarity with queried product is : 0.999893
Discount %:  0.0
--------------------------------------------------

417 : conditioner hair repair
Cosine Similarity with queried product is : 0.999886
Discount %:  0.0
--------------------------------------------------

15535 : satreetha shampoo
Cosine Similarity with queried product is : 0.999873
Discount %:  0.0
--------------------------------------------------

15427 : conditioning shampoo
Cosine Similarity with queried product is : 0.999873
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 22349 :
 product                                    keratin smooth conditioner
category                                              beauty_hygiene
sub_category                                                haircare
brand                                                       tresemme
type                                             shampoo_conditioner
description       deep condition your hair and protect them from...
sale_price                                                       165
discount_%                                                        25
negative                                                       0.036
neutral                                                        0.801
positive                                                       0.163
compound                                                      0.9794
cluster_label                                                      7
Name: 22349, dtype: object

Top 10 Similar products for "keratin smooth conditioner" are:
======================================================================

14101 : nourish replenish conditioner
Cosine Similarity with queried product is : 0.999971
Discount %:  25.0
--------------------------------------------------

20866 : climate protection conditioner
Cosine Similarity with queried product is : 0.999951
Discount %:  25.0
--------------------------------------------------

64 : aqua halo rejuvenating conditioner
Cosine Similarity with queried product is : 0.999933
Discount %:  25.0
--------------------------------------------------

26061 : hair fall defense conditioner
Cosine Similarity with queried product is : 0.999946
Discount %:  22.8663
--------------------------------------------------

924 : oil nourish conditioner
Cosine Similarity with queried product is : 0.999933
Discount %:  20.1005
--------------------------------------------------

15459 : color protect conditioner
Cosine Similarity with queried product is : 0.999933
Discount %:  20.0
--------------------------------------------------

1568 : herbal orange lemongrass hair conditioner
Cosine Similarity with queried product is : 0.999933
Discount %:  5.0
--------------------------------------------------

11821 : aloe hair conditioner
Cosine Similarity with queried product is : 0.999938
Discount %:  0.0
--------------------------------------------------

417 : conditioner hair repair
Cosine Similarity with queried product is : 0.999934
Discount %:  0.0
--------------------------------------------------

19337 : fructis long strong strengthening conditioner
Cosine Similarity with queried product is : 0.999933
Discount %:  0.0
--------------------------------------------------

The searched/Queried product is:
 22349 : keratin smooth conditioner

Top 10 Similar products for "keratin smooth conditioner" are:
======================================================================

924 : oil nourish conditioner
Cosine Similarity with queried product is : 0.999589
Discount %:  20.1005
--------------------------------------------------

15459 : color protect conditioner
Cosine Similarity with queried product is : 0.999606
Discount %:  20.0
--------------------------------------------------

5700 : advanced hair fall control conditioner
Cosine Similarity with queried product is : 0.999583
Discount %:  17.0
--------------------------------------------------

1568 : herbal orange lemongrass hair conditioner
Cosine Similarity with queried product is : 0.99967
Discount %:  5.0
--------------------------------------------------

18814 : silky smooth care shampoo conditioner
Cosine Similarity with queried product is : 0.999837
Discount %:  0.0
--------------------------------------------------

11821 : aloe hair conditioner
Cosine Similarity with queried product is : 0.99965
Discount %:  0.0
--------------------------------------------------

16553 : keratin smooth conditioner
Cosine Similarity with queried product is : 0.999649
Discount %:  0.0
--------------------------------------------------

9386 : conditioner lusciously thick long nourishing
Cosine Similarity with queried product is : 0.999632
Discount %:  0.0
--------------------------------------------------

25728 : hair conditioner green tea with aloevera
Cosine Similarity with queried product is : 0.999625
Discount %:  0.0
--------------------------------------------------

19337 : fructis long strong strengthening conditioner
Cosine Similarity with queried product is : 0.999583
Discount %:  0.0
--------------------------------------------------
```

In [40]:

```python
#Summarizing the results of all 3 methods
prod_index=22349
d=pd.DataFrame({'BOW':bow_22349['product'].values[1:],'TF-IDF':tfidf_22349['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_22349['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                          'props':[('color','black'),
                                                                                                                   ('font-weight', 'bold'),
                                                                                                                   ('text-align','left'),
                                                                                                                   ('font-size', '16px')]}])
```

Out[40]:

Similar products for : " keratin smooth conditioner"

	BOW	TF-IDF	TF-IDF_W2V
0	nourish replenish conditioner	nourish replenish conditioner	oil nourish conditioner
1	climate protection conditioner	climate protection conditioner	color protect conditioner
2	aqua halo rejuvenating conditioner	aqua halo rejuvenating conditioner	advanced hair fall control conditioner
3	hair fall defense conditioner	hair fall defense conditioner	herbal orange lemongrass hair conditioner
4	oil nourish conditioner	oil nourish conditioner	silky smooth care shampoo conditioner
5	color protect conditioner	color protect conditioner	aloe hair conditioner
6	aloe hair conditioner	herbal orange lemongrass hair conditioner	keratin smooth conditioner
7	conditioner hair repair	aloe hair conditioner	conditioner lusciously thick long nourishing
8	satreetha shampoo	conditioner hair repair	hair conditioner green tea with aloevera
9	conditioning shampoo	fructis long strong strengthening conditioner	fructis long strong strengthening conditioner

In [41]:

```python
# checking for prod index 789 , top 10 similar items
bow_789=bag_of_words_product_with_discount(789,11)
tfidf_789=tfidf_product(789,11)
tfidf_w2v_789=tfidf_w2v_product(789,11)
```

```
The searched\Queried product is:
 789 : tomato salsa dip enjoy with nacho chips

Top 10 Similar products for "tomato salsa dip enjoy with nacho chips" are:
======================================================================

11994 : yummy dip out
Cosine Similarity with queried product is : 0.999797
Discount %:  20.0
--------------------------------------------------

22526 : pizza foundue dip cheese
Cosine Similarity with queried product is : 0.999779
Discount %:  20.0
--------------------------------------------------

5431 : yogurt dip cilantro jalapeno
Cosine Similarity with queried product is : 0.999778
Discount %:  20.0
--------------------------------------------------

4439 : dip hot cheese
Cosine Similarity with queried product is : 0.999764
Discount %:  20.0
--------------------------------------------------

3005 : very creamy salsa spicy chipotle
Cosine Similarity with queried product is : 0.999761
Discount %:  20.0
--------------------------------------------------

5516 : ranch mint herb
Cosine Similarity with queried product is : 0.999758
Discount %:  20.0
--------------------------------------------------

347 : chunky salsa mild
Cosine Similarity with queried product is : 0.999765
Discount %:  15.0
--------------------------------------------------

19954 : chunky salsa hot
Cosine Similarity with queried product is : 0.999765
Discount %:  15.0
--------------------------------------------------

22781 : soy with chilli
Cosine Similarity with queried product is : 0.999763
Discount %:  0.0
--------------------------------------------------

7745 : sriracha dip sauce
Cosine Similarity with queried product is : 0.999761
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 789 :
 product                       tomato salsa dip enjoy with nacho chips
category                                              gourmet_worldfood
sub_category                                        sauces_spreads_dips
brand                                                          habanero
type                                                 hummus_cheese_salsadip
description      traditional mexican style family gathering hav...
sale_price                                                        157.5
discount_%                                                           10
negative                                                              0
neutral                                                           0.857
positive                                                          0.143
compound                                                         0.7184
cluster_label                                                         7
Name: 789, dtype: object

Top 10 Similar products for "tomato salsa dip enjoy with nacho chips" are:
======================================================================

15737 : all natural salsa peri peri salsa
Cosine Similarity with queried product is : 0.999932
Discount %:  20.0
--------------------------------------------------

12246 : very cheesy salsa salsa con queso
Cosine Similarity with queried product is : 0.999929
Discount %:  20.0
--------------------------------------------------

3005 : very creamy salsa spicy chipotle
Cosine Similarity with queried product is : 0.999928
Discount %:  20.0
--------------------------------------------------

22526 : pizza foundue dip cheese
Cosine Similarity with queried product is : 0.999925
Discount %:  20.0
--------------------------------------------------

11994 : yummy dip out
Cosine Similarity with queried product is : 0.999925
Discount %:  20.0
--------------------------------------------------

5431 : yogurt dip cilantro jalapeno
Cosine Similarity with queried product is : 0.999925
Discount %:  20.0
--------------------------------------------------

8799 : all natural hummus spicy sriracha
Cosine Similarity with queried product is : 0.99992
Discount %:  20.0
--------------------------------------------------

16573 : all natural hummus fresh hummus
Cosine Similarity with queried product is : 0.999919
Discount %:  20.0
--------------------------------------------------

5516 : ranch mint herb
Cosine Similarity with queried product is : 0.999918
Discount %:  20.0
--------------------------------------------------

3854 : all natural hummus roasted garlic
Cosine Similarity with queried product is : 0.999917
Discount %:  20.0
--------------------------------------------------

The searched/Queried product is:
 789 : tomato salsa dip enjoy with nacho chips

Top 10 Similar products for "tomato salsa dip enjoy with nacho chips" are:
======================================================================

26463 : very cheesy dip pepper jack
Cosine Similarity with queried product is : 0.999784
Discount %:  20.0
--------------------------------------------------

3005 : very creamy salsa spicy chipotle
Cosine Similarity with queried product is : 0.999781
Discount %:  20.0
--------------------------------------------------

12246 : very cheesy salsa salsa con queso
Cosine Similarity with queried product is : 0.999761
Discount %:  20.0
--------------------------------------------------

8799 : all natural hummus spicy sriracha
Cosine Similarity with queried product is : 0.999754
Discount %:  20.0
--------------------------------------------------

22526 : pizza foundue dip cheese
Cosine Similarity with queried product is : 0.999751
Discount %:  20.0
--------------------------------------------------

11994 : yummy dip out
Cosine Similarity with queried product is : 0.99975
Discount %:  20.0
--------------------------------------------------

3854 : all natural hummus roasted garlic
Cosine Similarity with queried product is : 0.999726
Discount %:  20.0
--------------------------------------------------

15737 : all natural salsa peri peri salsa
Cosine Similarity with queried product is : 0.999701
Discount %:  20.0
--------------------------------------------------

4439 : dip hot cheese
Cosine Similarity with queried product is : 0.999694
Discount %:  20.0
--------------------------------------------------

4848 : potato crisp chips hot spicy
Cosine Similarity with queried product is : 0.999692
Discount %:  0.0
--------------------------------------------------
```

In [42]:

```python
#Summarizing the results of all 3 methods
prod_index=789
d=pd.DataFrame({'BOW':bow_789['product'].values[1:],'TF-IDF':tfidf_789['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_789['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                          'props':[('color','black'),
                                                                                                                   ('font-weight', 'bold'),
                                                                                                                   ('text-align','left'),
                                                                                                                   ('font-size', '16px')]}])
```

Out[42]:

Similar products for : " tomato salsa dip enjoy with nacho chips"

	BOW	TF-IDF	TF-IDF_W2V
0	yummy dip out	all natural salsa peri peri salsa	very cheesy dip pepper jack
1	pizza foundue dip cheese	very cheesy salsa salsa con queso	very creamy salsa spicy chipotle
2	yogurt dip cilantro jalapeno	very creamy salsa spicy chipotle	very cheesy salsa salsa con queso
3	dip hot cheese	pizza foundue dip cheese	all natural hummus spicy sriracha
4	very creamy salsa spicy chipotle	yummy dip out	pizza foundue dip cheese
5	ranch mint herb	yogurt dip cilantro jalapeno	yummy dip out
6	chunky salsa mild	all natural hummus spicy sriracha	all natural hummus roasted garlic
7	chunky salsa hot	all natural hummus fresh hummus	all natural salsa peri peri salsa
8	soy with chilli	ranch mint herb	dip hot cheese
9	sriracha dip sauce	all natural hummus roasted garlic	potato crisp chips hot spicy

In [43]:

```python
# checking for prod index 6786 , top 10 similar items
bow_6786=bag_of_words_product_with_discount(6786,11)
tfidf_6786=tfidf_product(6786,11)
tfidf_w2v_6786=tfidf_w2v_product(6786,11)
```

```
The searched\Queried product is:
 6786 : dressing thousand island

Top 10 Similar products for "dressing thousand island" are:
======================================================================

14832 : thousand island dressing lite
Cosine Similarity with queried product is : 0.999992
Discount %:  0.0
--------------------------------------------------

8328 : french dressing
Cosine Similarity with queried product is : 0.999977
Discount %:  0.0
--------------------------------------------------

488 : dressing creamy caesar
Cosine Similarity with queried product is : 0.999972
Discount %:  0.0
--------------------------------------------------

4733 : italian dressing less fat
Cosine Similarity with queried product is : 0.999965
Discount %:  0.0
--------------------------------------------------

3757 : mayonnaise
Cosine Similarity with queried product is : 0.999955
Discount %:  0.0
--------------------------------------------------

9203 : tulsi honey
Cosine Similarity with queried product is : 0.999954
Discount %:  0.0
--------------------------------------------------

1427 : nutella ready
Cosine Similarity with queried product is : 0.999953
Discount %:  0.0
--------------------------------------------------

22517 : red pepper sauce
Cosine Similarity with queried product is : 0.999953
Discount %:  0.0
--------------------------------------------------

18661 : syrup pancake
Cosine Similarity with queried product is : 0.999952
Discount %:  0.0
--------------------------------------------------

16290 : mayonnaise mayolite
Cosine Similarity with queried product is : 0.99995
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 6786 :
 product                                       dressing thousand island
category                                              gourmet_worldfood
sub_category                                        sauces_spreads_dips
brand                                                    americangarden
type                                                    saladdressings
description      american garden thousand island takes you cook...
sale_price                                                          270
discount_%                                                            0
negative                                                              0
neutral                                                           0.893
positive                                                          0.107
compound                                                           0.34
cluster_label                                                         6
Name: 6786, dtype: object

Top 10 Similar products for "dressing thousand island" are:
======================================================================

6304 : original recipe italian herb mayo
Cosine Similarity with queried product is : 0.999971
Discount %:  30.0
--------------------------------------------------

14832 : thousand island dressing lite
Cosine Similarity with queried product is : 0.999997
Discount %:  0.0
--------------------------------------------------

488 : dressing creamy caesar
Cosine Similarity with queried product is : 0.99999
Discount %:  0.0
--------------------------------------------------

4733 : italian dressing less fat
Cosine Similarity with queried product is : 0.999989
Discount %:  0.0
--------------------------------------------------

8328 : french dressing
Cosine Similarity with queried product is : 0.999989
Discount %:  0.0
--------------------------------------------------

772 : peanut butter chunky
Cosine Similarity with queried product is : 0.999972
Discount %:  0.0
--------------------------------------------------

20889 : barbeque sauce original
Cosine Similarity with queried product is : 0.999971
Discount %:  0.0
--------------------------------------------------

18661 : syrup pancake
Cosine Similarity with queried product is : 0.999971
Discount %:  0.0
--------------------------------------------------

22517 : red pepper sauce
Cosine Similarity with queried product is : 0.999969
Discount %:  0.0
--------------------------------------------------

9203 : tulsi honey
Cosine Similarity with queried product is : 0.999969
Discount %:  0.0
--------------------------------------------------

The searched/Queried product is:
 6786 : dressing thousand island

Top 10 Similar products for "dressing thousand island" are:
======================================================================

7531 : olive oil extra virgin
Cosine Similarity with queried product is : 0.99986
Discount %:  58.6429
--------------------------------------------------

25645 : san remo pasta disano olive oil american garden pasta sauce
Cosine Similarity with queried product is : 0.999861
Discount %:  32.3486
--------------------------------------------------

3957 : spanish extra virgin olive oil
Cosine Similarity with queried product is : 0.999863
Discount %:  30.0
--------------------------------------------------

11204 : cashew nuts whole natural premium king size
Cosine Similarity with queried product is : 0.999861
Discount %:  24.726
--------------------------------------------------

10647 : chinese five spice powder shaker
Cosine Similarity with queried product is : 0.999863
Discount %:  10.0
--------------------------------------------------

14832 : thousand island dressing lite
Cosine Similarity with queried product is : 0.999979
Discount %:  0.0
--------------------------------------------------

4733 : italian dressing less fat
Cosine Similarity with queried product is : 0.999888
Discount %:  0.0
--------------------------------------------------

22173 : phantom hot tomato ketchup sauce made with world hottest ghost pepper
Cosine Similarity with queried product is : 0.999865
Discount %:  0.0
--------------------------------------------------

7620 : olive oil tomato and porcini mushroom pasta sauce with extra virgin
Cosine Similarity with queried product is : 0.999862
Discount %:  0.0
--------------------------------------------------

27007 : american garden peanut butter chunky jar
Cosine Similarity with queried product is : 0.999861
Discount %:  0.0
--------------------------------------------------
```

In [44]:

```python
#Summarizing the results of all 3 methods
prod_index=6786
d=pd.DataFrame({'BOW':bow_6786['product'].values[1:],'TF-IDF':tfidf_6786['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_6786['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                          'props':[('color','black'),
                                                                                                                   ('font-weight', 'bold'),
                                                                                                                   ('text-align','left'),
                                                                                                                   ('font-size', '16px')]}])
```

Out[44]:

Similar products for : " dressing thousand island"

	BOW	TF-IDF	TF-IDF_W2V
0	thousand island dressing lite	original recipe italian herb mayo	olive oil extra virgin
1	french dressing	thousand island dressing lite	san remo pasta disano olive oil american garden pasta sauce
2	dressing creamy caesar	dressing creamy caesar	spanish extra virgin olive oil
3	italian dressing less fat	italian dressing less fat	cashew nuts whole natural premium king size
4	mayonnaise	french dressing	chinese five spice powder shaker
5	tulsi honey	peanut butter chunky	thousand island dressing lite
6	nutella ready	barbeque sauce original	italian dressing less fat
7	red pepper sauce	syrup pancake	phantom hot tomato ketchup sauce made with world hottest ghost pepper
8	syrup pancake	red pepper sauce	olive oil tomato and porcini mushroom pasta sauce with extra virgin
9	mayonnaise mayolite	tulsi honey	american garden peanut butter chunky jar

In [47]:

```python
# checking for prod index 2607 , top 10 similar items
bow_2607=bag_of_words_product_with_discount(2607,11)
tfidf_2607=tfidf_product(2607,11)
tfidf_w2v_2607=tfidf_w2v_product(2607,11)
```

```
The searched\Queried product is:
 2607 : organic rava idli mix

Top 10 Similar products for "organic rava idli mix" are:
======================================================================

11541 : instant mix rava idli
Cosine Similarity with queried product is : 0.999615
Discount %:  10.0
--------------------------------------------------

3597 : organic masala rava idli ready mix
Cosine Similarity with queried product is : 0.999746
Discount %:  0.0
--------------------------------------------------

9853 : organic rice idli ready
Cosine Similarity with queried product is : 0.999591
Discount %:  0.0
--------------------------------------------------

23692 : ragi dosa mix
Cosine Similarity with queried product is : 0.999561
Discount %:  0.0
--------------------------------------------------

24212 : breakfast mix upma
Cosine Similarity with queried product is : 0.999561
Discount %:  0.0
--------------------------------------------------

20410 : mix millet pongal
Cosine Similarity with queried product is : 0.999556
Discount %:  0.0
--------------------------------------------------

17737 : multigrain thalipeeth mix
Cosine Similarity with queried product is : 0.999531
Discount %:  0.0
--------------------------------------------------

12169 : organic ragi dosa ready mix
Cosine Similarity with queried product is : 0.999526
Discount %:  0.0
--------------------------------------------------

7218 : organic rice dosa ready mix
Cosine Similarity with queried product is : 0.999525
Discount %:  0.0
--------------------------------------------------

16997 : ready mix sambar
Cosine Similarity with queried product is : 0.999525
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 2607 :
 product                                          organic rava idli mix
category                                              snacks_brandedfoods
sub_category                                             readytocook_eat
brand                                                    phaladapure_sure
type                                                   breakfast_snackmixes
description      rava idli variation the popular south indian b...
sale_price                                                           88
discount_%                                                            0
negative                                                           0.01
neutral                                                           0.826
positive                                                          0.164
compound                                                         0.9744
cluster_label                                                        14
Name: 2607, dtype: object

Top 10 Similar products for "organic rava idli mix" are:
======================================================================

7582 : breakfast mix oats pongal
Cosine Similarity with queried product is : 0.999768
Discount %:  28.8
--------------------------------------------------

26671 : breakfast mix little millet upma
Cosine Similarity with queried product is : 0.999763
Discount %:  28.8
--------------------------------------------------

2994 : breakfast mix little millet pongal
Cosine Similarity with queried product is : 0.999762
Discount %:  28.8
--------------------------------------------------

8288 : specialty idli dosa batter
Cosine Similarity with queried product is : 0.999765
Discount %:  18.0952
--------------------------------------------------

3597 : organic masala rava idli ready mix
Cosine Similarity with queried product is : 0.999847
Discount %:  0.0
--------------------------------------------------

9853 : organic rice idli ready
Cosine Similarity with queried product is : 0.999784
Discount %:  0.0
--------------------------------------------------

23692 : ragi dosa mix
Cosine Similarity with queried product is : 0.99977
Discount %:  0.0
--------------------------------------------------

24212 : breakfast mix upma
Cosine Similarity with queried product is : 0.999768
Discount %:  0.0
--------------------------------------------------

20410 : mix millet pongal
Cosine Similarity with queried product is : 0.999764
Discount %:  0.0
--------------------------------------------------

7218 : organic rice dosa ready mix
Cosine Similarity with queried product is : 0.999759
Discount %:  0.0
--------------------------------------------------

The searched/Queried product is:
 2607 : organic rava idli mix

Top 10 Similar products for "organic rava idli mix" are:
======================================================================

168 : organic dosa mix jowar
Cosine Similarity with queried product is : 0.999716
Discount %:  0.0
--------------------------------------------------

3597 : organic masala rava idli ready mix
Cosine Similarity with queried product is : 0.999403
Discount %:  0.0
--------------------------------------------------

12169 : organic ragi dosa ready mix
Cosine Similarity with queried product is : 0.999401
Discount %:  0.0
--------------------------------------------------

10347 : organic mix dal
Cosine Similarity with queried product is : 0.999389
Discount %:  0.0
--------------------------------------------------

16854 : organic rice mix flaxseed
Cosine Similarity with queried product is : 0.999361
Discount %:  0.0
--------------------------------------------------

20190 : organic dosa mix mixed millet
Cosine Similarity with queried product is : 0.999343
Discount %:  0.0
--------------------------------------------------

7218 : organic rice dosa ready mix
Cosine Similarity with queried product is : 0.99928
Discount %:  0.0
--------------------------------------------------

467 : organic rasam ready mix long pepper
Cosine Similarity with queried product is : 0.999071
Discount %:  0.0
--------------------------------------------------

4598 : organic dosa mix little millet
Cosine Similarity with queried product is : 0.999024
Discount %:  0.0
--------------------------------------------------

20562 : organic pongal mix foxtail millet
Cosine Similarity with queried product is : 0.999016
Discount %:  0.0
--------------------------------------------------
```

In [48]:

```python
#Summarizing the results of all 3 methods
prod_index=2607
d=pd.DataFrame({'BOW':bow_2607['product'].values[1:],'TF-IDF':tfidf_2607['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_2607['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                                                          'props':[('color','black'),
                                                                                                                   ('font-weight', 'bold'),
                                                                                                                   ('text-align','left'),
                                                                                                                   ('font-size', '16px')]}])
```

Out[48]:

Similar products for : " organic rava idli mix"

	BOW	TF-IDF	TF-IDF_W2V
0	instant mix rava idli	breakfast mix oats pongal	organic dosa mix jowar
1	organic masala rava idli ready mix	breakfast mix little millet upma	organic masala rava idli ready mix
2	organic rice idli ready	breakfast mix little millet pongal	organic ragi dosa ready mix
3	ragi dosa mix	specialty idli dosa batter	organic mix dal
4	breakfast mix upma	organic masala rava idli ready mix	organic rice mix flaxseed
5	mix millet pongal	organic rice idli ready	organic dosa mix mixed millet
6	multigrain thalipeeth mix	ragi dosa mix	organic rice dosa ready mix
7	organic ragi dosa ready mix	breakfast mix upma	organic rasam ready mix long pepper
8	organic rice dosa ready mix	mix millet pongal	organic dosa mix little millet
9	ready mix sambar	organic rice dosa ready mix	organic pongal mix foxtail millet

In [49]:

```
# checking for prod_index 19098 , top 10 similar items
bow_19098=bag_of_words_product_with_discount(19098,11)
tfidf_19098=tfidf_product(19098,11)
tfidf_w2v_19098=tfidf_w2v_product(19098,11)
```

```
The searched\Queried product is:
 19098 : millet muesli crunchy nutty delight

Top 10 Similar products for "millet muesli crunchy nutty delight" are:
======================================================================

16084 : millet diet muesli
Cosine Similarity with queried product is : 0.999921
Discount %:  30.0
--------------------------------------------------

14627 : oat crispies classic
Cosine Similarity with queried product is : 0.999793
Discount %:  30.0
--------------------------------------------------

13794 : hazelnut
Cosine Similarity with queried product is : 0.99979
Discount %:  23.0
--------------------------------------------------

25836 : instant oats
Cosine Similarity with queried product is : 0.999794
Discount %:  19.598
--------------------------------------------------

24266 : cashew
Cosine Similarity with queried product is : 0.999792
Discount %:  7.6923
--------------------------------------------------

21534 : rolled oats
Cosine Similarity with queried product is : 0.99981
Discount %:  5.0
--------------------------------------------------

18419 : oats flour
Cosine Similarity with queried product is : 0.999805
Discount %:  0.0
--------------------------------------------------

13380 : peanut butter crunchy
Cosine Similarity with queried product is : 0.999793
Discount %:  0.0
--------------------------------------------------

27060 : healthy snacks nutty apricot fibre
Cosine Similarity with queried product is : 0.999791
Discount %:  0.0
--------------------------------------------------

21005 : moments
Cosine Similarity with queried product is : 0.999791
Discount %:  0.0
--------------------------------------------------

The searched\Queried product is:
 19098 :
 product                         millet muesli crunchy nutty delight
category                                          gourmet_worldfood
sub_category                                      cereals_breakfast
brand                                                  milletmantra
type                                               muesli_ricecakes
description      crunchy nutty delight muesli millet based mues...
sale_price                                                      168
discount_%                                                       30
negative                                                      0.158
neutral                                                       0.645
positive                                                      0.197
compound                                                      0.743
cluster_label                                                     7
Name: 19098, dtype: object

Top 10 Similar products for "millet muesli crunchy nutty delight" are:
======================================================================

16084 : millet diet muesli
Cosine Similarity with queried product is : 0.999973
Discount %:  30.0
--------------------------------------------------

14627 : oat crispies classic
Cosine Similarity with queried product is : 0.999872
Discount %:  30.0
--------------------------------------------------

8201 : super oats rolled oats
Cosine Similarity with queried product is : 0.999877
Discount %:  22.0
--------------------------------------------------

2474 : teff grain vegan non gmo improves gut health
Cosine Similarity with queried product is : 0.999895
Discount %:  20.0
--------------------------------------------------

25836 : instant oats
Cosine Similarity with queried product is : 0.999885
Discount %:  19.598
--------------------------------------------------

10975 : whole instant oats gluten free
Cosine Similarity with queried product is : 0.999895
Discount %:  10.2564
--------------------------------------------------

21534 : rolled oats
Cosine Similarity with queried product is : 0.999897
Discount %:  5.0
--------------------------------------------------

27060 : healthy snacks nutty apricot fibre
Cosine Similarity with queried product is : 0.999898
Discount %:  0.0
--------------------------------------------------

18419 : oats flour
Cosine Similarity with queried product is : 0.999893
Discount %:  0.0
--------------------------------------------------

13380 : peanut butter crunchy
Cosine Similarity with queried product is : 0.999872
Discount %:  0.0
--------------------------------------------------

The searched/Queried product is:
 19098 : millet muesli crunchy nutty delight

Top 10 Similar products for "millet muesli crunchy nutty delight" are:
======================================================================

23236 : assorted crispy treats
Cosine Similarity with queried product is : 0.999682
Discount %:  30.0601
--------------------------------------------------

15249 : popcorn nutty tuxedo chocolate
Cosine Similarity with queried product is : 0.999718
Discount %:  15.0
--------------------------------------------------

4221 : flavoured cashew godambis creamy chocolate
Cosine Similarity with queried product is : 0.999683
Discount %:  15.0
--------------------------------------------------

5467 : unsweetened crunchy peanut butter
Cosine Similarity with queried product is : 0.999712
Discount %:  10.0
--------------------------------------------------

27040 : chocolate peanut butter crunchy
Cosine Similarity with queried product is : 0.999686
Discount %:  9.7506
--------------------------------------------------

19187 : organic peanut butter creamy sweet salty
Cosine Similarity with queried product is : 0.999711
Discount %:  7.5
--------------------------------------------------

9526 : crunchy organic peanut butter
Cosine Similarity with queried product is : 0.999691
Discount %:  5.1282
--------------------------------------------------

27060 : healthy snacks nutty apricot fibre
Cosine Similarity with queried product is : 0.999759
Discount %:  0.0
--------------------------------------------------

14608 : healthy snacks nutty apricot fibre
Cosine Similarity with queried product is : 0.999726
Discount %:  0.0
--------------------------------------------------

25120 : breakfast cereal crunchy granola with nutty honey
Cosine Similarity with queried product is : 0.99972
Discount %:  0.0
--------------------------------------------------
```

In [50]:

```
#Summarizing the results of all 3 methods
prod_index=19098
d=pd.DataFrame({'BOW':bow_19098['product'].values[1:],'TF-IDF':tfidf_19098['product'].values[1:], \
                'TF-IDF_W2V':tfidf_w2v_19098['product'].values[1:]})
# setting caption(title) to dataframe
d.style.set_caption('Similar products for : '+'" '+data['product'].loc[prod_index]+'"').set_table_styles([{'selector':'caption',
                                                                  'props':[('color','black'),
                                                                           ('font-weight', 'bold'),
                                                                           ('text-align','left'),
                                                                           ('font-size', '16px')]}])
```

Out[50]:

Similar products for : " millet muesli crunchy nutty delight"

	BOW	TF-IDF	TF-IDF_W2V
0	millet diet muesli	millet diet muesli	assorted crispy treats
1	oat crispies classic	oat crispies classic	popcorn nutty tuxedo chocolate
2	hazelnut	super oats rolled oats	flavoured cashew godambis creamy chocolate
3	instant oats	teff grain vegan non gmo improves gut health	unsweetened crunchy peanut butter
4	cashew	instant oats	chocolate peanut butter crunchy
5	rolled oats	whole instant oats gluten free	organic peanut butter creamy sweet salty
6	oats flour	rolled oats	crunchy organic peanut butter
7	peanut butter crunchy	healthy snacks nutty apricot fibre	healthy snacks nutty apricot fibre
8	healthy snacks nutty apricot fibre	oats flour	healthy snacks nutty apricot fibre
9	moments	peanut butter crunchy	breakfast cereal crunchy granola with nutty honey

</u>Observations

- From the above result summary TF-IDF weighted W2V results are better compare to other two methods

In [ ]: