

```
In [51]: import pandas as pd
import numpy as np
from sentence_transformers import SentenceTransformer, util
import torch
```

```
In [3]: amz = pd.read_csv("amz_com-ecommerce_sample.csv", encoding= 'unicode_escape')
fk = pd.read_csv("flipkart_com-ecommerce_sample.csv", encoding= 'unicode_escape')
```

```
In [4]: amz.head()
```

Out[4]:

product_category_tree		pid	retail_price	discounted_price	image	is_FK_Advantage_product	description	product_rating	overall_rating	brand	product_specification
["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2FF9KEDEFGF		982	438	["http://img5a.flixcart.com/image/short/u/4/a/...	False	Key Features of Alisha Solid Women's Cycling S...	No rating available	No rating available	Alisha	{"product_specification":{"key"=>"N...
["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY		32143	29121	["http://img6a.flixcart.com/image/sofa-bed/j/f/...	False	FabHomeDecor Fabric Double Sofa Bed (Finish Co...	No rating available	No rating available	FabHomeDecor	{"product_specification":{"key"=>"...
["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUBJGZXE		991	551	["http://img5a.flixcart.com/image/shoe/7/z/z/r/...	False	Key Features of AW Bellies Sandals Wedges Heel...	No rating available	No rating available	AW	{"product_specification":{"key"=>"I...
["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2F6HUZMQ6SJ		694	325	["http://img5a.flixcart.com/image/short/6/2/h/...	False	Key Features of Alisha Solid Women's Cycling S...	No rating available	No rating available	Alisha	{"product_specification":{"key"=>"N...
["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMSYARJ5		208	258	["http://img5a.flixcart.com/image/pet-shampoo/...	False	Specifications of Sicons All Purpose Arnica Do...	No rating available	No rating available	Sicons	{"product_specification":{"key"=>"P...

```
In [5]: fk.head()
```

Out[5]:

	uniq_id	crawl_timestamp	product_url	product_name	product_category_tree	pid	retail_price	discounted_price	
0	c2d766ca982eca8304150849735ffe9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2FF9KEDEFGF	999.0	379.0	["http://img5a.flixcart.com/image/short/u/4/a/...
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedecor-fabric-do...	FabHomeDecor Fabric Double Sofa Bed	["Furniture >> Living Room Furniture >> Sofa B...	SBEEH3QGU7MFYJFY	32157.0	22646.0	["http://img6a.flixcart.com/image/sofa-bed/j/f/...
2	f449ec65dcbc041b6ae5e6a32717d01b	2016-03-25 22:59:23 +0000	http://www.flipkart.com/aw-bellies/p/itmeh4grg...	AW Bellies	["Footwear >> Women's Footwear >> Ballerinas >...	SHOEH4GRSUBJGZXE	999.0	499.0	["http://img5a.flixcart.com/image/shoe/7/z/z/r/...
3	0973b37acd0c664e3de26e97e5571454	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-solid-women-s-c...	Alisha Solid Women's Cycling Shorts	["Clothing >> Women's Clothing >> Lingerie, Sl...	SRTEH2F6HUZMQ6SJ	699.0	267.0	["http://img5a.flixcart.com/image/short/6/2/h/...
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	2016-03-25 22:59:23 +0000	http://www.flipkart.com/sicons-all-purpose-arn...	Sicons All Purpose Arnica Dog Shampoo	["Pet Supplies >> Grooming >> Skin & Coat Care...	PSOEH3ZYDMSYARJ5	220.0	210.0	["http://img5a.flixcart.com/image/pet-shampoo/...

```
In [44]: amz.dtypes
```

```
Out[44]: uniq_id          object
crawl_timestamp         object
product_url            object
product_name           object
product_category_tree  object
pid                   object
retail_price            int64
discounted_price        int64
image                  object
is_FK_Advantage_product bool
description             object
product_rating          object
overall_rating          object
brand                  object
product_specifications object
dtype: object
```

```
In [45]: fk.dtypes
```

```
Out[45]: uniq_id          object
crawl_timestamp         object
product_url            object
product_name           object
product_category_tree  object
pid                   object
retail_price            float64
discounted_price        float64
image                  object
is_FK_Advantage_product bool
description             object
product_rating          object
overall_rating          object
brand                  object
product_specifications object
dtype: object
```

```
In [6]: amz.shape, fk.shape
```

```
Out[6]: ((20000, 15), (20000, 15))
```

```
In [7]: amz.isna().sum()
```

```
Out[7]: uniq_id          0
crawl_timestamp         0
product_url            0
product_name           0
product_category_tree  0
pid                   0
retail_price            0
discounted_price        0
image                  3
is_FK_Advantage_product 0
description             2
product_rating          0
overall_rating          0
brand                  5864
product_specifications 14
dtype: int64
```

```
In [8]: fk.isna().sum()
```

```
Out[8]: uniq_id                0
crawl_timestamp              0
product_url                 0
product_name                0
product_category_tree       0
pid                        0
retail_price                78
discounted_price            78
image                      3
is_FK_Advantage_product    0
description                 2
product_rating              0
overall_rating              0
brand                     5864
product_specifications      14
dtype: int64
```

```
In [9]: dfa = amz[['product_name', 'retail_price', 'discounted_price']]
dff = fk[['product_name', 'retail_price', 'discounted_price']]
```

```
In [10]: dfa.head()
```

```
Out[10]:
```

	product_name	retail_price	discounted_price
0	Alisha Solid Women's Cycling Shorts	982	438
1	FabHomeDecor Fabric Double Sofa Bed	32143	29121
2	AW Bellies	991	551
3	Alisha Solid Women's Cycling Shorts	694	325
4	Sicons All Purpose Arnica Dog Shampoo	208	258

```
In [12]: dff.head()
```

```
Out[12]:
```

	product_name	retail_price	discounted_price
0	Alisha Solid Women's Cycling Shorts	999.0	379.0
1	FabHomeDecor Fabric Double Sofa Bed	32157.0	22646.0
2	AW Bellies	999.0	499.0
3	Alisha Solid Women's Cycling Shorts	699.0	267.0
4	Sicons All Purpose Arnica Dog Shampoo	220.0	210.0

Sentence Transformer Reference - <https://www.sbert.net/> (<https://www.sbert.net/>)

```
In [16]: model = SentenceTransformer('all-MiniLM-L6-v2')
```

```
Downloading: 100% 1.18k/1.18k [00:00<00:00, 78.3kB/s]
Downloading: 100% 190/190 [00:00<00:00, 13.6kB/s]
Downloading: 100% 10.6k/10.6k [00:00<00:00, 707kB/s]
Downloading: 100% 612/612 [00:00<00:00, 43.7kB/s]
Downloading: 100% 116/116 [00:00<00:00, 7.72kB/s]
Downloading: 100% 39.3k/39.3k [00:00<00:00, 50.2kB/s]
Downloading: 100% 90.9M/90.9M [00:13<00:00, 6.84MB/s]
Downloading: 100% 53.0/53.0 [00:00<00:00, 3.53kB/s]
Downloading: 100% 112/112 [00:00<00:00, 7.46kB/s]
Downloading: 100% 466k/466k [00:01<00:00, 362kB/s]
Downloading: 100% 350/350 [00:00<00:00, 25.0kB/s]
Downloading: 100% 13.2k/13.2k [00:00<00:00, 38.0kB/s]
Downloading: 100% 232k/232k [00:01<00:00, 201kB/s]
Downloading: 100% 349/349 [00:00<00:00, 24.9kB/s]
```

```
In [17]: dfa_en = model.encode(dfa.product_name,convert_to_tensor=True)
dff_en = model.encode(dff.product_name,convert_to_tensor=True)
cosine_sim = util.cos_sim(dfa_en,dff_en)
```

```
In [18]: sim_products_index = []
for i in range(20000):
    m = torch.topk(cosine_sim[i],k=5,largest=True,sorted=True).indices
    for ind in m:
        if ind != i:
            sim_products_index.append(ind)
            break
```

```
In [20]: sim_products_index = [x.item() for x in sim_products_index]
```

```
In [33]: fk_d = []
        amz_d = []

        for ind, prod_name in enumerate(dfa.product_name):
            fk_d.append(dff.iloc[sim_products_index[ind]])
            amz_d.append(dfa.iloc[ind])
```

```
In [34]: fk_d = pd.DataFrame(fk_d)
        fk_d.columns = ['Product name in Flipkart', 'Retail Price in Flipkart', 'Discounted Price in Flipkart']

        amz_d = pd.DataFrame(amz_d)
        amz_d.columns = ['Product name in Amazon', 'Retail Price in Amazon', 'Discounted Price in Amazon']
```

```
In [38]: fk_d.head(10)
```

Out[38]:

	Product name in Flipkart	Retail Price in Flipkart	Discounted Price in Flipkart
3	Alisha Solid Women's Cycling Shorts	699.0	267.0
16	FabHomeDecor Fabric Double Sofa Bed	32157.0	22646.0
17691	i-Sports Bellies	899.0	399.0
0	Alisha Solid Women's Cycling Shorts	999.0	379.0
20	Sicons Conditioning Conditoner Dog Shampoo	110.0	100.0
5251	Aaishwarya Glim Glam Crystal Crystal Alloy Nec...	1450.0	599.0
0	Alisha Solid Women's Cycling Shorts	999.0	379.0
16	FabHomeDecor Fabric Double Sofa Bed	32157.0	22646.0
15764	Provogue Corporate Casuals	2485.0	2485.0
3	Alisha Solid Women's Cycling Shorts	699.0	267.0

```
In [40]: amz_d.dtypes
```

Out[40]: Product name in Amazon object  
Retail Price in Amazon int64  
Discounted Price in Amazon int64  
dtype: object

```
In [49]: fk_d.dtypes
```

Out[49]: Product name in Flipkart object  
Retail Price in Flipkart float64  
Discounted Price in Flipkart float64  
dtype: object

```
In [52]: amz_d['Retail Price in Amazon'] = amz_d['Retail Price in Amazon'].astype(np.float64)
        amz_d['Discounted Price in Amazon'] = amz_d['Discounted Price in Amazon'].astype(np.float64)
```

```
In [53]: amz_d.head(10)
```

Out[53]:

	Product name in Amazon	Retail Price in Amazon	Discounted Price in Amazon
0	Alisha Solid Women's Cycling Shorts	982.0	438.0
1	FabHomeDecor Fabric Double Sofa Bed	32143.0	29121.0
2	AW Bellies	991.0	551.0
3	Alisha Solid Women's Cycling Shorts	694.0	325.0
4	Sicons All Purpose Arnica Dog Shampoo	208.0	258.0
5	Eternal Gandhi Super Series Crystal Paper Weig...	427.0	473.0
6	Alisha Solid Women's Cycling Shorts	1198.0	602.0
7	FabHomeDecor Fabric Double Sofa Bed	32137.0	28664.0
8	dilli bazaaar Bellies, Corporate Casuals, Casuals	682.0	385.0
9	Alisha Solid Women's Cycling Shorts	1197.0	542.0

```
In [54]: amz_d.shape, fk_d.shape
```

Out[54]: ((20000, 3), (20000, 3))

```
In [56]: fk_d = fk_d.reset_index(drop=True)
```

```
In [57]: final_list = pd.concat([fk_d,amz_d],axis=1)
```

```
In [60]: final_list.head(10)
```

Out[60]:

	Product name in Flipkart	Retail Price in Flipkart	Discounted Price in Flipkart	Product name in Amazon	Retail Price in Amazon	Discounted Price in Amazon
0	Alisha Solid Women's Cycling Shorts	699.0	267.0	Alisha Solid Women's Cycling Shorts	982.0	438.0
1	FabHomeDecor Fabric Double Sofa Bed	32157.0	22646.0	FabHomeDecor Fabric Double Sofa Bed	32143.0	29121.0
2	i-Sports Bellies	899.0	399.0	AW Bellies	991.0	551.0
3	Alisha Solid Women's Cycling Shorts	999.0	379.0	Alisha Solid Women's Cycling Shorts	694.0	325.0
4	Sicons Conditioning Conditoner Dog Shampoo	110.0	100.0	Sicons All Purpose Arnica Dog Shampoo	208.0	258.0
5	Aaishwarya Glim Glam Crystal Crystal Alloy Nec...	1450.0	599.0	Eternal Gandhi Super Series Crystal Paper Weig...	427.0	473.0
6	Alisha Solid Women's Cycling Shorts	999.0	379.0	Alisha Solid Women's Cycling Shorts	1198.0	602.0
7	FabHomeDecor Fabric Double Sofa Bed	32157.0	22646.0	FabHomeDecor Fabric Double Sofa Bed	32137.0	28664.0
8	Provogue Corporate Casuals	2485.0	2485.0	dilli bazaaar Bellies, Corporate Casuals, Casuals	682.0	385.0
9	Alisha Solid Women's Cycling Shorts	699.0	267.0	Alisha Solid Women's Cycling Shorts	1197.0	542.0