



# Ecommerce Recommendation System

# Table of contents

01

Overview  
Project

02

Data Understanding &  
Preparation

03

Explanatory Data  
Analysis

04

Recommendation  
System



01

# Overview Project

# Myntra.com

Myntra adalah perusahaan e-commerce fashion besar di India yang berkantor pusat di Bengaluru, Karnataka, India. Perusahaan ini didirikan pada tahun 2007 untuk menjual barang hadiah yang dipersonalisasi. Pada Mei 2014, Myntra.com diakuisisi oleh Flipkart.



# Objectives



## Tujuan

Merekomendasikan produk dari brand tertentu berdasarkan preferensi brand yang disukai user, sehingga dapat mempermudah serta meningkatkan user experience.



## Strategi

Membangun sistem rekomendasi content based filtering, sehingga dapat mempermudah konsumen untuk membandingkan serta memilih produk/brand berdasarkan preferensi personal. Serta mampu merekomendasikan produk/brand baru dan populer.



## Problem Ideas

1. Persebaran produk/brand termahal dan termurah
2. Bagaimana persebaran harga?
3. Bagaimana persebaran brand yang ada?



## Goal

1. Melihat persebaran harga dan brand yang ada
2. Membuat sistem rekomendasi berdasarkan deskripsi produk dari setiap brand.





02

# Data Understanding & Preparation

# Data Understanding

## Information

Pada data, terdapat:

**12491 rows × 8 columns**

Dengan terdapat missing Value pada kolom primary color  
dan tidak terdapat data duplikat

Dengan kolom:

- ProductID
- ProductName
- Product Brand
- Gender
- Price (INR)
- NumImages
- Description
- PrimaryColor



# Data Preparation

ProductID	0
ProductName	0
ProductBrand	0
Gender	0
Price (INR)	0
NumImages	0
Description	0
PrimaryColor	894
dtype:	int64

Terdapat missing value  
pada kolom primary color



ProductID	0
ProductName	0
ProductBrand	0
Gender	0
Price (INR)	0
NumImages	0
Description	0
PrimaryColor	0
dtype:	int64

Missing Value primary color diubah  
dgn warna yang paling banyak muncul  
(blue)

	ProductID	ProductName	ProductBrand	Gender	Price (INR)	NumImages	Description	PrimaryColor
0	10017413	DKNY Unisex Black & Grey Printed Medium Trolle...	DKNY	Unisex	11745	7	Black and grey printed medium trolley bag, sec...	Black
1	10016283	EthnoVogue Women Beige & Grey Made to Measure ...	EthnoVogue	Women	5810	7	Beige & Grey made to measure kurta with churid...	Beige
2	10009781	SPYKAR Women Pink Alexa Super Skinny Fit High-...	SPYKAR	Women	899	7	Pink coloured wash 5-pocket high-rise cropped ...	Pink
3	10015921	Raymond Men Blue Self-Design Single-Breasted B...	Raymond	Men	5599	5	Blue self-design bandhgala suitBlue self-desig...	Blue
4	10017833	Parx Men Brown & Off-White Slim Fit Printed Ca...	Parx	Men	759	5	Brown and off-white printed casual shirt, has ...	White
...	...	...	...	...	...	...	...	...
12486	10262843	Pepe Jeans Men Black Hammock Slim Fit Low-Rise...	Pepe Jeans	Men	1299	7	Black dark wash 5-pocket low-rise jeans, clean...	Black
12487	10261721	Mochi Women Gold-Toned Solid Heels	Mochi	Women	1990	5	A pair of gold-toned open toe heels, has regul...	Gold
12488	10261607	612 league Girls Navy Blue & White Printed Reg...	612 league	Girls	602	4	Navy Blue and White printed mid-rise denim sho...	Blue
12489	10266621	Bvlgari Men Aqva Pour Homme Marine Eau de Toil...	Bvlgari	Men	8950	2	Bvlgari Men Aqva Pour Homme Marine Eau de Toil...	NaN
12490	10265199	Pepe Jeans Men Black & Grey Striped Polo Colla...	Pepe Jeans	Men	799	5	Black and grey striped T-shirt, has a polo col...	Black

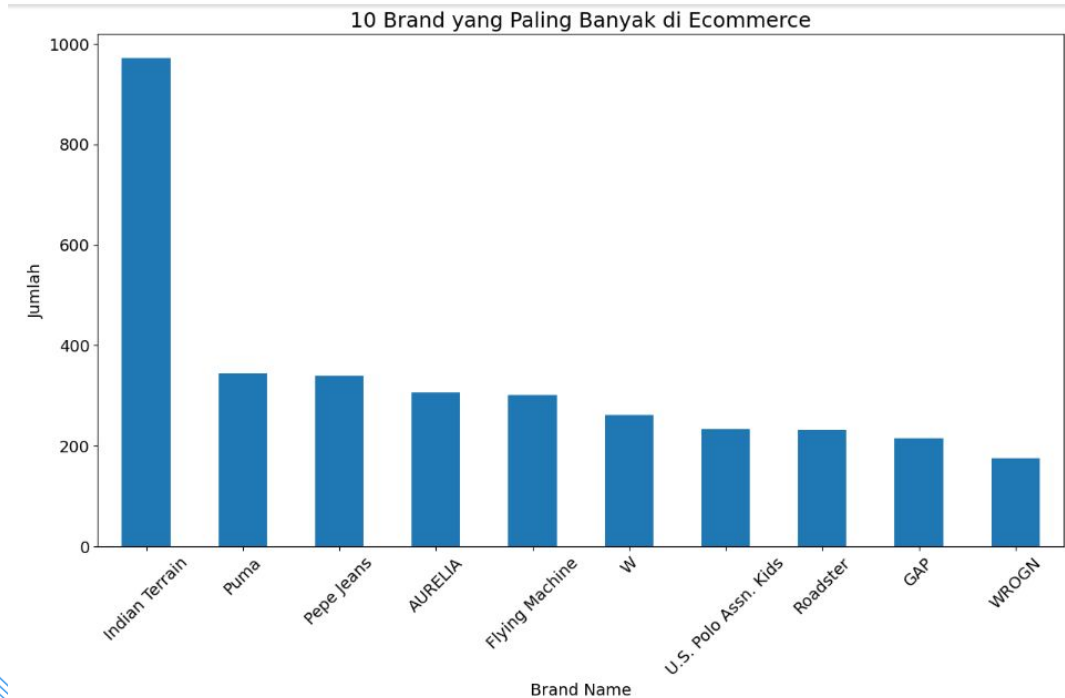
12491 rows × 8 columns

Pada proses ini, kolom yang digunakan adalah ProductID, ProductName, ProductBrand, Price (INR), dan Description.



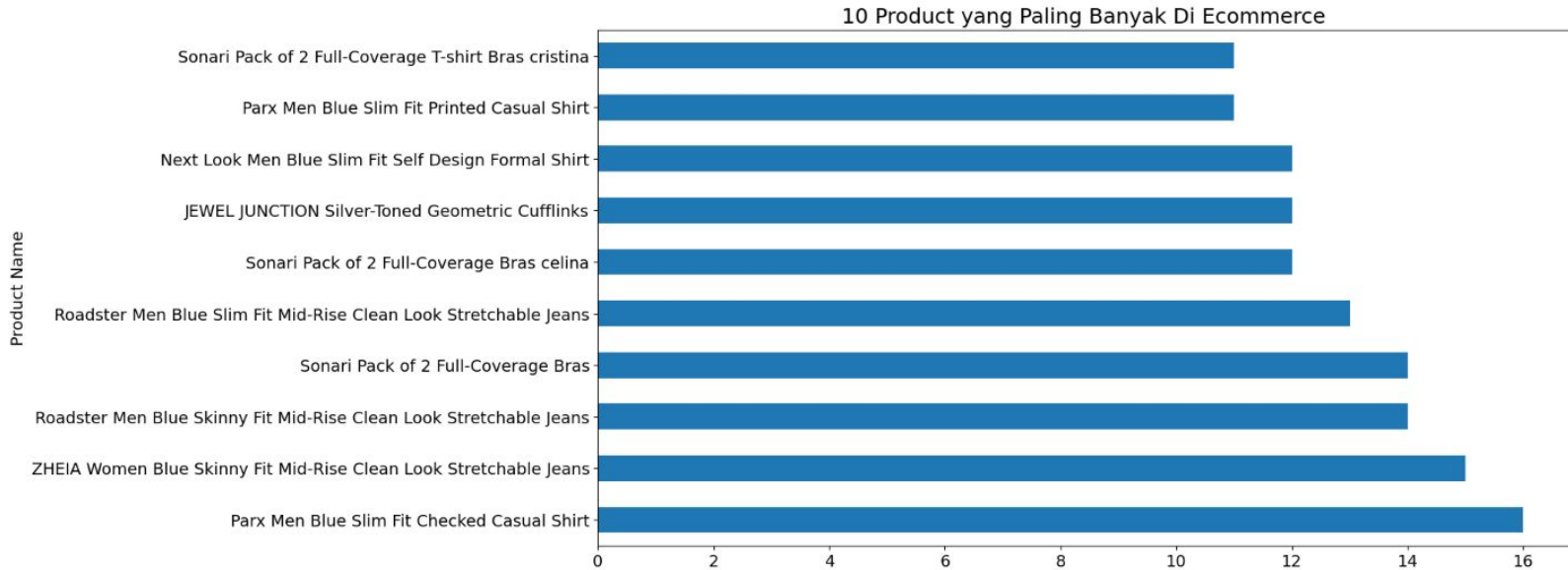
03

# Eksploratory Data Analysis



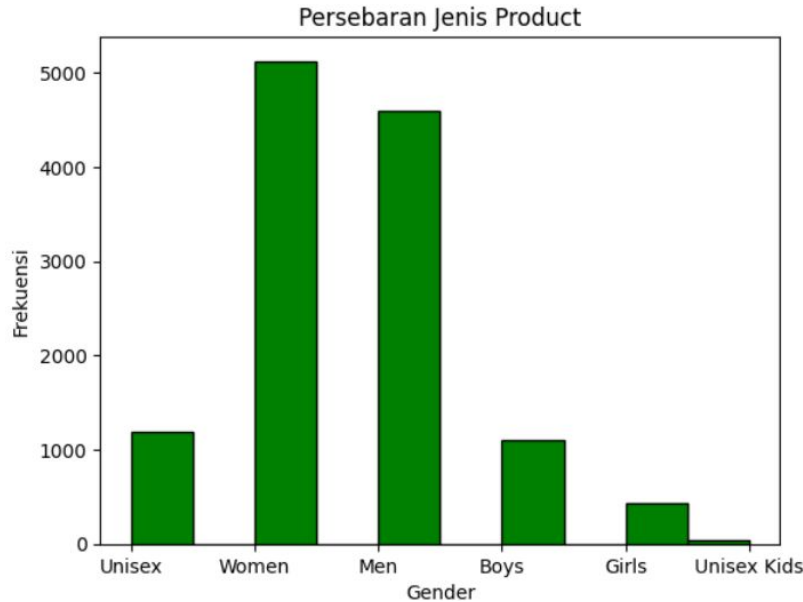
10 Brand terbanyak didominasi oleh brand-brand fashion.

Brand indian terrain merupakan brand yang paling diminati, maka akan lebih baik untuk brand ini selalu ready stock dan menjaga kualitas serta performanya.



## 10 Produk Paling Banyak Diminati di Ecommerce

Untuk 10 produk yang paling banyak diminati, dapat disediakan dalam berbagai ukuran yang ready stock. Sehingga user tidak perlu menunggu. Sedangkan, untuk produk yang masih kurang peminat, dapat memberlakukan promo atau product bundling, sehingga dapat meningkatkan daya beli produk/brand dan menghasilkan revenue.



Untuk persebaran section product, didominasi oleh section Women, disusul Men dan Unisex.

Kebanyakan pembeli di Ecommerce adalah wanita, sehingga untuk section women dan men dapat terus dijaga untuk stock nya, jika bisa ditambah variasi/style nya agar user tidak jenuh.

Dan untuk unisex kid dan girls, karena peminatnya sedikit, dapat diatur pendistribusiannya di ecommerce dan dapat dilakukan promo/bundling dengan produk2 section women atau men.



	Price (INR)
ProductBrand	
Garmin	63090.000000
MOVADO	49192.666667
Maserati	23200.000000
SEIKO	18438.888889
BERING	16941.800000
DKNY	14814.482759
Jacques Lemans	11100.000000
PRESTO	10981.076923
MASPAR	9499.000000
Calvin Klein	8508.384615

	Price (INR)
ProductName	
Garmin Unisex Blue Forerunner 945 Smartwatch 010-02063-71	63090.0
Garmin Unisex Black Forerunner 945 Smartwatch 010-02063-70	63090.0
MOVADO Men Black Analogue Watch 607243	58854.0
MOVADO Women Blue Analogue Watch 607268	56192.0
SEIKO Solar Women Silver Analogue watch SUP428P1	34500.0
MOVADO Men Blue Swiss-Made Analogue Watch 607197	32532.0
DKNY Unisex Purple Large Trolley Bag	31100.0
PRESTO Beige & Black Printed Handmade Anti-Skid Woolen Carpet	29998.0
SEIKO Solar Women Off White Analogue watch SUP432P1	29500.0
SEIKO Solar Men Black Analogue watch SSC723P1	29500.0

10 Brand dan Produk dengan harga tertinggi didominasi oleh brand/produk gadget/aksesoris.

brand/produk gadget dan aksesoris, kurang diminati oleh user. Dikarenakan harga yang mahal sehingga user lebih memilih untuk langsung ke toko dibanding online.

Ecommerce dapat membuat imbauan/campaign untuk meyakinkan user atau memberikan jaminan, bahwa membeli produk gadget/aksesoris di platform mendapatkan asuransi keamanan, baik jika barang itu hilang, terdapat scam ataupun rusak saat pengiriman.



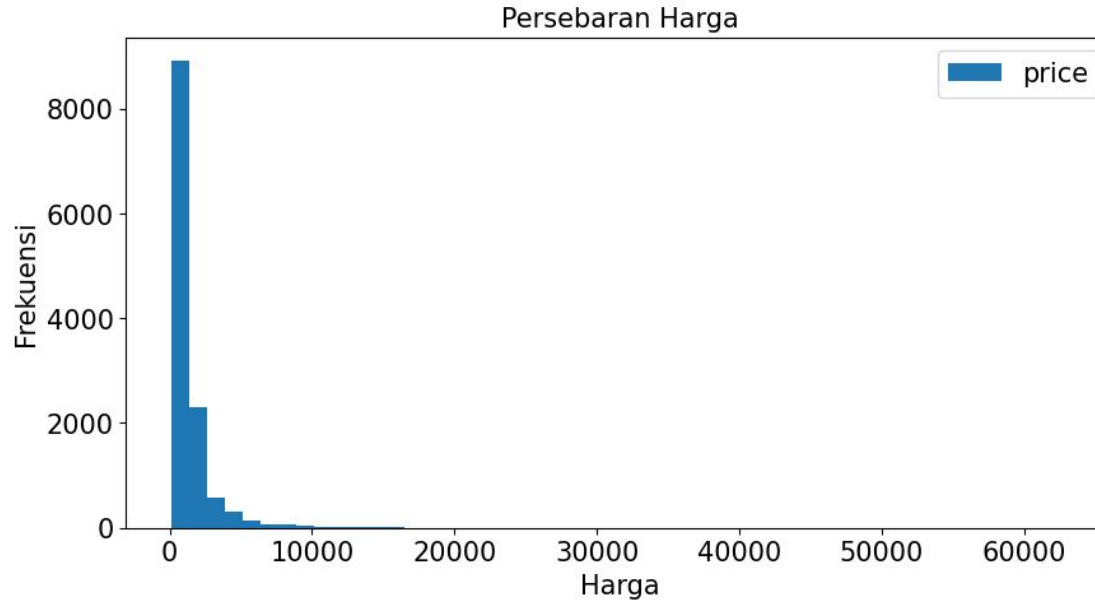
ProductBrand	Price (INR)
Envy	139.000000
Ponds	158.333333
Rosaline	291.000000
St. Ives	299.000000
Morpheme Remedies	299.666667
INCOLOR	322.761905
UNKNOWN by Ayesha	324.000000
BEARDO	330.500000
Foolzy	339.000000
Secret Temptation	344.666667

ProductName	Price (INR)
Organic Harvest Unisex Sulphate Free Skin Lightening Face Wash 50g	90.0
Organic Harvest Unisex Shine & Glow Sheet Mask 20gm	91.0
Organic Harvest Unisex Moisturising Sheet Face Mask 20g	91.0
Organic Harvest Unisex Anti-Wrinkle Sheet Face Mask 20g	91.0
Organic Harvest Unisex Anti-Acne Sheet Face Mask 20g	91.0
Ponds Men Daily Defence SPF 30 Face Creme 55 g	132.0
Ponds Super Light Gel Oil Free Moisturiser 73 g	134.0
Envy Men Absolute Perfume Deodorant Spray - 120 ml	139.0
Organic Harvest Unisex Sulphate Free Oil Control Face Wash 100g	150.0
Organic Harvest Unisex Sulphate Free Acne Control Face Wash 100g	150.0

Sedangkan, untuk 10 Brand dengan harga terendah, didominasi oleh brand/produk skincare.

Untuk brand skincare/bodycare, dapat fokus pada pembuatan campaign-campaign yang lebih menonjolkan manfaat dan keuntungan dari brand tersebut. Serta, menjaga loyalitas user, seperti memberikan promo kepada user yang sudah lama menggunakan suatu brand/membership. Dan dapat juga melakukan penawaran bundling

	Price (INR)	frekuensi
321	699	510
703	1299	464
965	1999	312
247	599	295
473	899	254
540	999	196
176	499	183
358	749	183
270	629	178
138	449	175
398	799	169
283	649	124
203	539	121
646	1199	119
103	399	118
335	719	115
802	1499	113
435	849	108
428	839	102
300	674	102



Distribusi harga didominasi dengan produk/brand di <3000 INR

Kebanyakan user, lebih memilih produk/brand yang masih affordable atau <3000 INR. yang berarti preferensi user terhadap harga lebih condong untuk membeli yang murah.



04

# Recommendation System

# Rekomendasi Sistem Metode CBF

```
#Text Preprocessing

clean_spcl = re.compile('[/(){}\\[\\]\\|@,;]')
clean_symbol = re.compile('[^0-9a-z #+_]')
stopword = set(stopwords.words('english'))

def clean_text(text):
    text = text.lower() # lowercase text
    text = clean_spcl.sub(' ', text)
    text = clean_symbol.sub('', text)
    text = ' '.join(word for word in text.split() if word not in stopwords)
    return text

#kolom tambahan untuk data description yang telah dibersihkan
fashion['desc_clean'] = fashion['Description'].apply(clean_text)
```

text preprocessing yang bertujuan agar data yang dipakai nanti dapat diproses menjadi angka dengan TF-IDF dan cosine similarity. Data yang akan dipakai untuk diproses menjadi angka adalah kolom 'description' karena agar bisa mendapatkan kesamaan pada pengaplikasiannya nanti.

```
def recommendations(ProductBrand, cos_sim = cos_sim):  
  
    recommended_brand = []  
  
    # Mengambil nama brand berdasarkan variabel indices  
    idx = indices[indices == ProductBrand].index[0]  
  
    # Membuat series berdasarkan skor kesamaan  
    score_series = pd.Series(cos_sim[idx]).sort_values(ascending = False)  
  
    # mengambil index dan dibuat 10 baris rekomendasi terbaik  
    top_10_indexes = list(score_series.iloc[1:11].index)  
  
    for i in top_10_indexes:  
        recommended_brand.append(list(fashion.index)[i])  
  
    return recommended_brand
```

```
recommendations('Mochi')
```

```
['Mochi',  
 'MSC',  
 'GNIST',  
 'Inc 5',  
 'GNIST',  
 'Metro',  
 'FOREVER 21',  
 'MSC',  
 'Metro',  
 'MSC']
```

Dengan rekomendasi sistem yang menggunakan metode content-based filtering dapat menghasilkan nama-nama brand yang memiliki kesamaan dari sisi deskripsi yang ada pada dataset, sehingga user mendapatkan rekomendasi sesuai prefensinya sendiri.

Contoh seperti di samping, ketika user mencari brand “Mochi” maka akan muncul juga produk/brand yang hampir serupa dengan brand “Mochi”.



# Rekomendasi Sistem Metode CF

```
fashion2[['id', 'ClothingID', 'Rating']]
```

	id	ClothingID	Rating
0	0	767	4
1	1	1080	5
2	2	1077	3
3	3	1049	5
4	4	847	5
...	...	...	...
23481	23481	1104	5
23482	23482	862	3
23483	23483	1104	3
23484	23484	1084	3
23485	23485	1104	5

23486 rows × 3 columns



```
reader = Reader()
data = Dataset.load_from_df(fashion2[['id', 'ClothingID', 'Rating']], reader)

# Train & Test
trainset, testset = train_test_split(data, test_size=0.20, random_state=42)
```

```
model_svd = SVD()
model_svd.fit(trainset)
```

```
<surprise.prediction_algorithms.matrix_factorization.SVD at 0x7d250c1d2920>
```

Pada metode CF ini, hanya beberapa kolom yang digunakan, yaitu kolom id, ClothingID, dan rating.

Setelah itu melakukan train dan test dari variabel yang sudah dibuat.

```
▶ model_svd = SVD()  
model_svd.fit(trainset)
```

```
🔗 <surprise.prediction_algorithms.matrix_factorization.SVD at 0x7d250c1d2920>
```

Setelah melakukan train dan test, selanjutnya ke pemodelan. Di sini, menggunakan SVD.

## ✓ Matrix Factorization with SVD

```
▶ #SVD  
prediction = model_svd.fit(trainset).test(testset)  
  
# Prediksi  
prediction
```

```
[Prediction(uid=21213, iid=1081, r_ui=4.0, est=4.274095617719967, details={'was_impossible': False}),  
Prediction(uid=191, iid=895, r_ui=2.0, est=4.344549661986209, details={'was_impossible': False}),  
Prediction(uid=12901, iid=1078, r_ui=3.0, est=4.163154372071556, details={'was_impossible': False}),  
Prediction(uid=13233, iid=1078, r_ui=4.0, est=4.163154372071556, details={'was_impossible': False}),  
Prediction(uid=12629, iid=820, r_ui=4.0, est=4.22356984796429, details={'was_impossible': False}),  
Prediction(uid=4485, iid=850, r_ui=2.0, est=4.389627085052248, details={'was_impossible': False}),  
Prediction(uid=9783, iid=824, r_ui=4.0, est=4.229571165702466, details={'was_impossible': False}),  
Prediction(uid=3924, iid=460, r_ui=2.0, est=4.317943906240201, details={'was_impossible': False}),  
Prediction(uid=9940, iid=873, r_ui=4.0, est=4.319590006125364, details={'was_impossible': False}),  
...
```



```
#Testing
```

```
recom_svd = model_svd.predict(uid='21213', iid='1081')
```

```
recom_svd
```

```
Prediction(uid='21213', iid='1081', r_ui=None, est=4.196987438790718, details={'was_impossible': False})
```

dari hasil prediksi di atas, memungkinkan untuk merekomendasikan produk dengan id=1081 kepada user id 21213

# Rekomendasi Sistem Metode K-NN

	between 2 and 4	under 2	upper 4	Rating
Class Name				
Intimates	1	0	0	4
Dresses	0	0	1	5
Dresses	1	0	0	3
Pants	0	0	1	5
Blouses	0	0	1	5
...	...	...	...	...
Dresses	0	0	1	5
Knits	1	0	0	3
Dresses	1	0	0	3
Dresses	1	0	0	3
Dresses	0	0	1	5

Di samping adalah hasil data yang telah melalui preprocessing.

Pada metode ini, menggunakan 2 kolom, yaitu class name dan rating.

```
def product_recommended(class_name=list(fashion2['Class Name'].value_counts().index)):
    class_list_name = []
    item_id = fashion2[fashion2['Class Name']==class_name].index
    item_id = item_id[0]
    for new_id in idlist[item_id]:
        class_list_name.append(fashion2.loc[new_id]['Class Name'])
    return class_list_name
```

```
product_recommended('Jeans')
```

```
['Knits',
 'Lounge',
 'Knits',
 'Dresses',
 'Dresses',
 'Knits',
 'Dresses',
 'Shorts',
 'Dresses',
 'Knits']
```

Dengan metode K-NN, dapat menampilkan produk yang berdasarkan tingkat kesamaan produk satu dengan lainnya dari pencarian produk yang paling sering dicari oleh user.

# Rekomendasi Sistem Berdasarkan Brand

```
@interact
def recommend_product_by_brand(brand_name = list(fashion['ProductBrand'].value_counts().index)):
    recommend = fashion[fashion['ProductBrand']==brand_name][['ProductName', 'Price (INR)']]
    recommend = recommend.sort_values(by = 'Price (INR)', ascending=True)
    return recommend.head(10)
```

brand\_name

ProductName Price (INR)



4186	Indian Terrain Boys Blue Printed Round Neck T-...	314
4667	Indian Terrain Boys Yellow Printed Round Neck ...	314
3972	Indian Terrain Boys Purple Printed Round Neck ...	314
4270	Indian Terrain Boys White Printed Round Neck T...	314
4703	Indian Terrain Boys Red Printed Round Neck T-s...	314
4095	Indian Terrain Boys Turquoise Blue Printed Rou...	314
4796	Indian Terrain Boys Coral Printed Round Neck T...	314
4468	Indian Terrain Boys Blue Printed Round Neck T-...	314
4381	Indian Terrain Boys Rust Brown Printed Round N...	314
4284	Indian Terrain Boys Yellow Printed Round Neck ...	359

Dari rekomendasi sistem yang sudah dibuat, menunjukan jika user mencari nama brand, maka akan muncul produk-produk dari brand tersebut dengan variasi produk dan harga tertentu.

# Content-Based Filtering

Model dengan content-based filtering memiliki beberapa keuntungan, yaitu:

- Memahami konten atau karakteristik dari item yang direkomendasikan. Sehingga dapat memberikan rekomendasi sesuai preferensi user.
- Dapat menangani item baru yang belum pernah dinilai pengguna.
- Tidak memerlukan data pengguna lainnya
- Mampu mengatasi masalah Cold Start

Kekurangan:

- Tidak dapat merekomendasikan item yang di luar preferensi user
- Kompleksitas pengembangan model fitur

# Collaborative Filtering with SVD

Model dengan collaborative filtering memiliki beberapa keuntungan, yaitu:

- Mampu menangkap pola hubungan kompleks antara user dan item berdasarkan riwayat interaksi user dengan item.
- Dapat menangani data yang sangat sparse
- Dapat diterapkan pada dataset yang besar

Kekurangan:

- Masih menghadapi kesulitan ketika mengatasi masalah Cold Start pada item baru yang belum memiliki sejarah interaksi.
- Keterbatasan dalam Menangani Data Non-Linearity dan Dinamika
- Keterbatasan dalam Menangani Sparsity Ekstrem
- Keterbatasan dalam Menangani Data Baru

# K-NN

Model dengan K-NN memiliki beberapa keuntungan, yaitu:

- sederhana dan mudah dimengerti. Karena, menghitung kesamaan antara item atau pengguna berdasarkan tetangga terdekat.
- Didasarkan pada preferensi dan perilaku langsung dari tetangga terdekat
- Dapat Menangani Perubahan Dinamika

Kekurangan:

- Sensitive terhadap Outliers
- Pengaruh pada Performa ketika K Kecil atau Besar
- Perhitungan kemiripan antar item atau pengguna dapat memakan waktu, terutama pada dataset yang besar





# Thanks!

**Ada Pertanyaan?**

[zakiyfuadi@outlook.com](mailto:zakiyfuadi@outlook.com)

+62 87873450949

<https://www.linkedin.com/in/fu-adi-zakiy-0a7a35193/>

# Reference

- <https://github.com/bayuzen19/Recommendation-System>
- <https://www.geeksforgeeks.org/recommendation-system-in-python/>
- <https://medium.com/data-folks-indonesia/recommendation-system-dengan-python-content-based-filtering-part-2-222a8c365add>
- <https://medium.com/data-folks-indonesia/recommendation-system-dengan-python-definisi-part-1-71154dc3f700>
- <https://www.kaggle.com/datasets/shivamb/fashion-clothing-products-catalog/data>
- <https://www.youtube.com/watch?v=7GaAp0XkLpM>
- <https://medium.com/data-folks-indonesia/recommendation-system-dengan-python-collaborative-filtering-part-3-219931e08fd8>