

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import sklearn
import sklearn.metrics as sm
from sklearn import linear_model
import matplotlib.pyplot as plt
```

```
df = pd.read_csv(r'/content/sample_data/newfash.csv')
```

```
# explore data
df.shape
```

```
(14310, 10)
```

```
df.head()
```

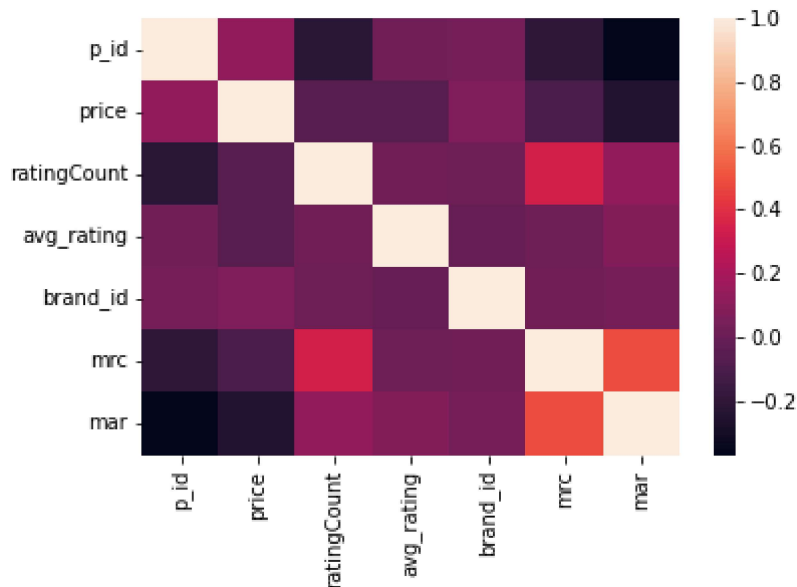
	brand_name	p_id	Product_name	price	color	ratingCount	avg_rating	brand_:
0	DUPATTA BAZAAR	1518329	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	1321.0	4.548827	242
1	ROADSTER	5829334	Roadster Women Mustard Yellow Solid	1199.0	Mustard	5462.0	4.313255	750

```
df.describe()
```

```

n_id      price  ratingCount  avg_rating  brand_id      mrc
matrix = df.corr().round(2)
sns.heatmap(matrix, annot=False)
plt.show()

```



```

print(df.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14310 entries, 0 to 14309
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype  
---  -
 0   brand_name      14310 non-null  object 
 1   p_id            14310 non-null  int64  
 2   Product_name    14310 non-null  object 
 3   price           14307 non-null  float64
 4   color           14304 non-null  object 
 5   ratingCount     6579 non-null   float64
 6   avg_rating      6579 non-null   float64
 7   brand_id        14307 non-null   float64
 8   mrc             14307 non-null   float64
 9   mar             14307 non-null   float64
dtypes: float64(6), int64(1), object(3)
memory usage: 1.1+ MB
None

```

```

print(df.dtypes)

brand_name      object
p_id            int64
Product_name    object
price           float64
color           object
ratingCount     float64

```

```

avg_rating    float64
brand_id      float64
mrc           float64
mar           float64
dtype: object

```

```
print(df.describe())
```

```

count      p_id      price  ratingCount  avg_rating  brand_id \
count  1.431000e+04  14307.000000  6579.000000  6579.000000  14307.000000
mean    1.569106e+07  2964.289788   184.530324    4.101139   531.449011
std     3.153509e+06  2564.254328   782.614622    0.475679   292.620391
min     7.016600e+04   169.000000    1.000000    1.000000    0.000000
25%    1.413618e+07  1599.000000    9.000000    3.888889   278.000000
50%    1.638216e+07  2200.000000   23.000000    4.180451   551.000000
75%    1.808344e+07  3495.000000   80.000000    4.392789   783.000000
max    1.941576e+07  47999.000000  21274.000000  5.000000  1020.000000

count      mrc      mar
count  14307.000000  14307.000000
mean     84.848243    1.885855
std    181.182462    1.219170
min      0.000000    0.000000
25%     2.500000    0.842857
50%    14.720721    1.967998
75%     99.142857    2.928773
max    6076.000000    4.833333

```

```

sorted_fashionset = df.sort_values(by='brand_id')
print(sorted_fashionset)

```

```

brand_name  p_id \
11379      BIBA  12414526
459      SASSAFRAS  7687291
8430      NEUDIS  13847264
1410      IVES  8375271
12278  VENI VIDI VICI  5293906
...      ...      ...
9423      ZOLA  14547528
13017      ZOLA  15033334
46      HOUSE OF PATAUDI  17022192
5063      YASH GALLERY  16919756
6727  HRX BY HRITHIK ROSHAN  13695238

Product_name  price  color \
11379  Biba Women Yellow & Pink Floral Print Maxi Fla...  3599.0  Yellow
459    SASSAFRAS Blue Denim Midi A-Line Pure Cotton S...  2199.0   Blue
8430  NEUDIS Women Olive-Green Self Design Flared Mi...  1199.0  Olive
1410                Ives Olive Green Shirt Style Top  1099.0  Olive
12278  Veni Vidi Vici Women Red Solid Trumpet Skirt  1170.0   Red
...      ...      ...
9423    ZOLA Women Black Skinny Fit Light Fade Jeans  2850.0  Black
13017  ZOLA Women Maroon Ethnic Motifs Embroidered Ch...  4650.0  Maroon
46    House of Pataudi Girls Green Pure Cotton Print...   NaN   NaN

```

5063	YASH GALLERY Brown	NaN	NaN
6727	HRX By Hrithik Roshan Women Wild Lime	NaN	NaN

	ratingCount	avg_rating	brand_id	mrc	mar
11379	NaN	NaN	0.0	14.720721	1.837091
459	577.0	4.214905	0.0	545.210526	3.588146
8430	32.0	4.468750	0.0	11.500000	1.563181
1410	5648.0	4.017882	0.0	1416.625000	2.992932
12278	NaN	NaN	0.0	691.000000	3.011795
...
9423	11.0	3.909091	1020.0	10.522727	1.852303
13017	133.0	4.248120	1020.0	10.522727	1.852303
46	NaN	NaN	NaN	NaN	NaN
5063	NaN	NaN	NaN	NaN	NaN
6727	NaN	NaN	NaN	NaN	NaN

[14310 rows x 10 columns]

sorted_fashionset.head()

	brand_name	p_id	Product_name	price	color	ratingCount	avg_rating	brar
11379	BIBA	12414526	Biba Women Yellow & Pink Floral Print Maxi Fla...	3599.0	Yellow	NaN	NaN	
459	SASSAFRAS	7687291	SASSAFRAS Blue Denim Midi A-Line Pure Cotton s	2199.0	Blue	577.0	4.214905	

df

	brand_name	p_id	Product_name	price	color	ratingCount	avg_rating	br
	DUPATTA	1518329	Dupatta Bazaar White Embroidered	899.0	White	1321.0	4.548827	

return the value

```
result = df.loc[:, 'price']
```

Print the result

```
print(result)
```

```
0      899.0
```

```
1     1199.0
```

```
2     5799.0
```

```
3     1499.0
```

```
4     1999.0
```

```
...
```

```
14305    3999.0
```

```
14306    2050.0
```

```
14307    1659.0
```

```
14308    2399.0
```

```
14309    2599.0
```

```
Name: price, Length: 14310, dtype: float64
```

```
fs=df
```

```
fs['ratingCount'] = fs['ratingCount'].fillna(fs['mrc'])
```

```
fs
```

```

brand_name    p_id Product_name    price    color    ratingCount    avg_rating    br
Dupatta
fs['avg_rating'] = fs['avg_rating'].fillna(fs['mar'])
fs

```

	brand_name	p_id	Product_name	price	color	ratingCount	avg_rating	br
0	DUPATTA BAZAAR	1518329	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	1321.000000	4.548827	
1	ROADSTER	5829334	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0	Mustard	5462.000000	4.313255	
2	INDDUS	10340119	Inddus Peach- Coloured & Beige Unstitched Dress...	5799.0	Peach	145.000000	4.068966	
3	SASSAFRAS	10856380	SASSAFRAS Women Black Parallel Trousers	1499.0	Black	9124.000000	4.147523	

```

print(fs.info())

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14310 entries, 0 to 14309
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   brand_name      14310 non-null  object
1   p_id            14310 non-null  int64
2   Product_name    14310 non-null  object
3   price           14307 non-null  float64
4   color           14304 non-null  object
5   ratingCount     14307 non-null  float64
6   avg_rating      14307 non-null  float64
7   brand_id        14307 non-null  float64
8   mrc             14307 non-null  float64
9   mar             14307 non-null  float64
dtypes: float64(6), int64(1), object(3)
memory usage: 1.1+ MB
None

```

```
fs=fs.fillna(0)
```

fs

	brand_name	p_id	Product_name	price	color	ratingCount	avg_rating	br
0	DUPATTA BAZAAR	1518329	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	1321.000000	4.548827	
1	ROADSTER	5829334	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0	Mustard	5462.000000	4.313255	
2	INDDUS	10340119	Inddus Peach- Coloured & Beige Unstitched Dress...	5799.0	Peach	145.000000	4.068966	
3	SASSAFRAS	10856380	SASSAFRAS Women Black Parallel Trousers	1499.0	Black	9124.000000	4.147523	

fs.to_csv(r'FashionSet.csv')