

```
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
```

```
#importing file to a dataframe
df1 = pd.read_csv(r'/content/sample_data/brandcsv.csv')
```

```
df1.head()
```

	brand_id	brand
0	1	513
1	2	109F
2	3	20Dresses
3	4	250 Designs
4	5	3Pin

```
df1.rename(columns = {'brand':'brand_name'}, inplace=True)
```

```
df1.head()
```

	brand_id	brand_name
0	1	513
1	2	109F
2	3	20Dresses
3	4	250 Designs
4	5	3Pin

```
df2 = pd.read_csv(r'/content/sample_data/fashionset.csv')
```

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

```
df2.rename(columns = {'name':'Product_name'}, inplace=True)
```

```
df = df2.merge(df1, on='brand_name', how='left')
print(df)
```

	p_id	Product_name	price \
0	1518329.0	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0
1	5829334.0	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0
2	10340119.0	Inddus Peach-Coloured & Beige Unstitched Dress...	5799.0
3	10856380.0	SASSAFRAS Women Black Parallel Trousers	1499.0
4	12384822.0	Kotty Women Black Wide Leg High-Rise Clean Loo...	1999.0
...	...	...	...
14323	17029604.0	The Chennai Silks Pink & Silver-Toned Floral Z...	3999.0
14324	17600212.0	Kinder Kids Girls Blue & Green Printed Foil Pr...	2050.0
14325	18159266.0	KLOTTHE Women Green & Black Floral Printed Pal...	1659.0
14326	18921114.0	InWeave Women Red Printed A-Line Skirt	2399.0
14327	19361058.0	BoStreet Women Navy Blue Tapered Fit Trousers	2599.0

	colour	brand_name	ratingCount	avg_rating	brand_id
0	White	Dupatta Bazaar	1321.0	4.548827	242.0
1	Mustard	Roadster	5462.0	4.313255	750.0
2	Peach	Inddus	145.0	4.068966	389.0
3	Black	SASSAFRAS	9124.0	4.147523	NaN
4	Black	Kotty	12260.0	4.078467	482.0
...	...	...	...	...	...
14323	Pink	The Chennai Silks	NaN	NaN	880.0
14324	Blue	Kinder Kids	NaN	NaN	471.0
14325	Green	KLOTTHE	NaN	NaN	NaN
14326	Red	InWeave	NaN	NaN	NaN
14327	Navy Blue	BoStreet	NaN	NaN	NaN

```
[14328 rows x 8 columns]
```

```
dfs.head()
```

```
brand_name;"p_id";"Product_name";"price";"color";"ratingCount";"avg_rating";"brand_i
```

0	DUPATTA BAZAAR;1518329;"Dupatta
1	ROADSTER;5829334;"Roadster Wome
2	INDDUS;10340119;"Inddus Peach-C
3	SASSAFRAS;10856380;"SASSAFRAS Wo
4	KOTTY;12384822;"Kotty Women Blar

```
# explore data
```

```
df.shape
```

```
(14328, 9)
```

```
df.to_csv(r'E:\big data\Assignment 2\pandasmergedfile.csv')  
df.to_excel("E:\big data\Assignment 2\pandasmergedfile.xlsx")
```

```
#importing file to a dataframe  
df = pd.read_csv(r"/content/sample_data/pandasmergedfile.csv")
```

```
#check for missing values  
df.isna().sum()
```

```
Unnamed: 0      0  
p_id           17  
Product_name    18  
price          18  
colour         21  
brand_name     23  
ratingCount    7747  
avg_rating     7747  
brand_id       6230  
dtype: int64
```

```
dffashion=df.fillna(0)  
dffashion
```

Unnamed: 0      p\_id   Product\_name   price   colour   brand\_name   ratingCount   avg

dffashion.describe()

	Unnamed: 0	p_id	price	ratingCount	avg_rating	brand_id
count	14328.000000	1.432800e+04	14328.000000	14328.000000	14328.000000	14328.000000
mean	7163.500000	1.567267e+07	2960.444654	84.733319	1.883736	296.402638
std	4136.281664	3.197613e+06	2564.554060	538.208854	2.069144	344.037170
min	0.000000	0.000000e+00	0.000000	0.000000	0.000000	0.000000
25%	3581.750000	1.412893e+07	1598.000000	0.000000	0.000000	0.000000
50%	7163.500000	1.637990e+07	2199.000000	0.000000	0.000000	125.000000
75%	10745.250000	1.807618e+07	3490.000000	19.000000	4.140000	583.000000
max	14327.000000	1.941576e+07	47999.000000	21274.000000	5.000000	1017.000000
Parallel						

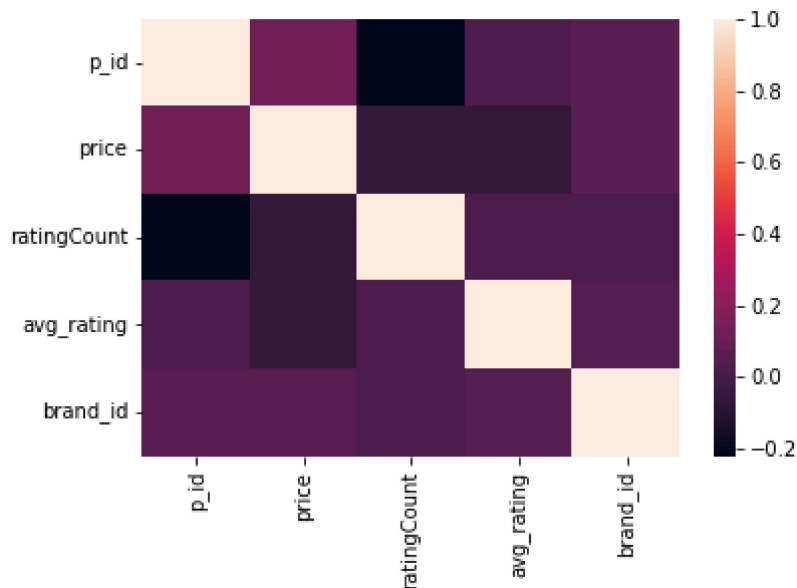
df.drop('Unnamed: 0', axis=1)

	p_id	Product_name	price	colour	brand_name	ratingCount	avg_ra
0	1518329.0	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	Dupatta Bazaar	1321.0	4.54
1	5829334.0	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0	Mustard	Roadster	5462.0	4.31
2	10340119.0	Inddus Peach-Coloured & Beige Unstitched Dress...	5799.0	Peach	Inddus	145.0	4.06
3	10856380.0	SASSAFRAS Women Black Parallel Trousers	1499.0	Black	SASSAFRAS	9124.0	4.14
4	12384822.0	Kotty Women Black Wide Leg High-Rise Clean Loo...	1999.0	Black	Kotty	12260.0	4.07
...	...	...	...	...	...	...	...
14323	17029604.0	The Chennai Silks Pink & Silver-Toned Floral	3999.0	Pink	The Chennai Silks	NaN	

dfn=df.drop('Unnamed: 0', axis=1)  
dfn.shape

(14328, 8)

```
matrix = dfn.corr().round(2)
sns.heatmap(matrix, annot=False)
plt.show()
```



```
dfn.to_csv(r'file1.csv')
```

```
dfn = pd.read_csv(r'/content/sample_data/file1.csv')
dfmc = pd.read_csv(r'/content/sample_data/mc.csv')
dfmar = pd.read_csv(r'/content/sample_data/mar.csv')
```

```
temp = dfn.merge(dfmc, on='brand_id', how='left')
print(temp)
```

```

      Unnamed: 0      p_id \
0              0  1518329.0
1              1   5829334.0
2              2  10340119.0
3              3  10856380.0
4              4  12384822.0
...           ...      ...
14323          14323  17029604.0
14324          14324  17600212.0
14325          14325  18159266.0
14326          14326  18921114.0
14327          14327  19361058.0
```

```

      Product_name  price  colour \
0  Dupatta Bazaar White Embroidered Chiffon Dupatta   899.0    White
1  Roadster Women Mustard Yellow Solid Hooded Swe...  1199.0  Mustard
2  Inddus Peach-Coloured & Beige Unstitched Dress...   5799.0    Peach
3  SASSAFRAS Women Black Parallel Trousers          1499.0    Black
4  Kotty Women Black Wide Leg High-Rise Clean Loo...   1999.0    Black
```

```

...
14323 The Chennai Silks Pink & Silver-Toned Floral Z... 3999.0 Pink
14324 Kinder Kids Girls Blue & Green Printed Foil Pr... 2050.0 Blue
14325 KLOTTHE Women Green & Black Floral Printed Pal... 1659.0 Green
14326 InWeave Women Red Printed A-Line Skirt 2399.0 Red
14327 BoStreet Women Navy Blue Tapered Fit Trousers 2599.0 Navy Blue

```

```

      brand_name  ratingCount  avg_rating  brand_id  mc
0      Dupatta Bazaar      1321.0    4.548827    242.0  29.502924
1          Roadster      5462.0    4.313255    750.0  300.667630
2          Inddus       145.0    4.068966    389.0  154.011494
3      SASSAFRAS       9124.0    4.147523     NaN  1251.400000
4          Kotty      12260.0    4.078467    482.0  1154.666667

```

```

...
14323 The Chennai Silks      NaN      NaN    880.0    0.000000
14324   Kinder Kids      NaN      NaN    471.0    0.000000
14325     KLOTTHE      NaN      NaN     NaN  1251.400000
14326     InWeave      NaN      NaN     NaN  1251.400000
14327     BoStreet      NaN      NaN     NaN  1251.400000

```

[14328 rows x 10 columns]

```

fashionset = temp.merge(dfmar, on='brand_id', how='left')
print(fashionset)

```

```

      Unnamed: 0      p_id \
0              0    1518329.0
1              1    5829334.0
2              2   10340119.0
3              3   10856380.0
4              4   12384822.0
...
14323      14323   17029604.0
14324      14324   17600212.0
14325      14325   18159266.0
14326      14326   18921114.0
14327      14327   19361058.0

```

```

      Product_name  price  colour \
0  Dupatta Bazaar White Embroidered Chiffon Dupatta    899.0   White
1  Roadster Women Mustard Yellow Solid Hooded Swe...   1199.0  Mustard
2  Inddus Peach-Coloured & Beige Unstitched Dress...   5799.0   Peach
3  SASSAFRAS Women Black Parallel Trousers          1499.0   Black
4  Kotty Women Black Wide Leg High-Rise Clean Loo...   1999.0   Black
...
14323 The Chennai Silks Pink & Silver-Toned Floral Z...   3999.0   Pink
14324 Kinder Kids Girls Blue & Green Printed Foil Pr...   2050.0   Blue
14325 KLOTTHE Women Green & Black Floral Printed Pal...   1659.0   Green
14326 InWeave Women Red Printed A-Line Skirt         2399.0   Red
14327 BoStreet Women Navy Blue Tapered Fit Trousers   2599.0  Navy Blue

```

```

      brand_name  ratingCount  avg_rating  brand_id  mc \
0      Dupatta Bazaar      1321.0    4.548827    242.0  29.502924
1          Roadster      5462.0    4.313255    750.0  300.667630
2          Inddus       145.0    4.068966    389.0  154.011494

```

3	SASSAFRAS	9124.0	4.147523	NaN	1251.400000
4	Kotty	12260.0	4.078467	482.0	1154.666667
...	...	...	...	...	...
14323	The Chennai Silks	NaN	NaN	880.0	0.000000
14324	Kinder Kids	NaN	NaN	471.0	0.000000
14325	KLOTTHE	NaN	NaN	NaN	1251.400000
14326	InWeave	NaN	NaN	NaN	1251.400000
14327	BoStreet	NaN	NaN	NaN	1251.400000

	mar
0	1.967998
1	3.532504
2	2.836643
3	2.540307
4	1.738539
...	...
14323	0.000000
14324	0.000000
14325	2.540307
14326	2.540307
14327	2.540307

[14328 rows x 11 columns]

fashionset.shape

(14328, 10)

print(fashionset.info())

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 14328 entries, 0 to 14327
Data columns (total 10 columns):
#   Column          Non-Null Count  Dtype
---  -
0   p_id            14311 non-null  float64
1   Product_name    14310 non-null  object
2   price           14310 non-null  float64
3   colour          14307 non-null  object
4   brand_name      14305 non-null  object
5   ratingCount     6581 non-null   float64
6   avg_rating      6581 non-null   float64
7   brand_id        8098 non-null   float64
8   mc              14328 non-null  float64
9   mar             14328 non-null  float64
dtypes: float64(7), object(3)
memory usage: 1.2+ MB
None
```

print(fashionset.dtypes)

```

p_id          float64
Product_name  object
price         float64
colour        object
brand_name    object
ratingCount   float64
avg_rating    float64
brand_id      float64
mc            float64
mar           float64
dtype: object

```

```
print(fashionset.describe())
```

```

count  p_id      price  ratingCount  avg_rating  brand_id \
mean   1.431100e+04  2964.168484   184.479410    4.101226   524.432823
std    3.153525e+06  2564.014851   782.501137    0.475633   299.716776
min    7.016600e+04   169.000000    1.000000    1.000000    1.000000
25%    1.413618e+07  1599.000000    9.000000    3.888889   242.000000
50%    1.638217e+07  2200.000000   23.000000    4.180822   554.000000
75%    1.808452e+07  3495.000000   80.000000    4.392857   776.000000
max    1.941576e+07  47999.000000  21274.000000  5.000000  1017.000000

count      mc      mar
mean    596.992811    2.238611
std     585.776780    0.950364
min       0.000000    0.000000
25%     15.411765    1.953687
50%     295.304348    2.540307
75%    1251.400000    2.540307
max    1625.000000    4.833333

```

```

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

```

```

p_id      Product_name  price \
598   13158392.0  513 Women Black & Grey Woven-Design Kimono Shrug  1699.0
6326  15883214.0    513 Women Black & Grey Checked Poncho Sweater  2499.0
12894 16920884.0    109F Women Black Striped Relaxed Trousers  1599.0
7129  17743260.0    109F Brown Culotte Jumpsuit  1999.0
13559 13200776.0    109F Women Navy Blue Slim Fit Jeans  1899.0
...      ...      ...      ...
14314 18055840.0  tantkatha Black Front Closure Saree Blouse Wit...  1699.0
14316 19361072.0    BoStreet Women Green Solid Mom Fit Trousers  2599.0
14325 18159266.0  KLOTTHE Women Green & Black Floral Printed Pal...  1659.0
14326 18921114.0    InWeave Women Red Printed A-Line Skirt  2399.0
14327 19361058.0    BoStreet Women Navy Blue Tapered Fit Trousers  2599.0

colour  brand_name  ratingCount  avg_rating  brand_id  mc \
598    Black      513        136.0    4.573529      1.0   86.5
6326    Black      513         37.0    4.486486      1.0   86.5

```



12894	Black	109F	NaN	NaN	2.0	0.0
7129	Brown	109F	NaN	NaN	2.0	0.0
13559	Navy Blue	109F	NaN	NaN	2.0	0.0
...	...	...	...	...	...	...
14314	Black	tantkatha	NaN	NaN	NaN	1251.4
14316	Green	BoStreet	NaN	NaN	NaN	1251.4
14325	Green	KLOTTHE	NaN	NaN	NaN	1251.4
14326	Red	InWeave	NaN	NaN	NaN	1251.4
14327	Navy Blue	BoStreet	NaN	NaN	NaN	1251.4

	mar
598	4.530008
6326	4.530008
12894	0.000000
7129	0.000000
13559	0.000000
...	...
14314	2.540307
14316	2.540307
14325	2.540307
14326	2.540307
14327	2.540307

[14328 rows x 10 columns]

sorted\_fashionset.head()

fashionset

	Unnamed: 0	p_id	Product_name	price	colour	brand_name	ratingCount	avg
0	0	1518329.0	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	Dupatta Bazaar	1321.0	4
1	1	5829334.0	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0	Mustard	Roadster	5462.0	4
2	2	10340119.0	Inddus Peach-Coloured & Beige Unstitched Dress...	5799.0	Peach	Inddus	145.0	4
3	3	10856380.0	SASSAFRAS Women Black Parallel	1499.0	Black	SASSAFRAS	9124.0	4

```
fashionset=fashionset.drop('Unnamed: 0', axis=1)
fashionset.shape
```

```
(14328, 10)
```

```
fashionset.to_csv(r'fashionsetpandastrimmed.csv')
```

```
# return the value
result = fashionset.loc[:, 'price']
```

```
# Print the result
print(result)
```

```
0      899.0
1     1199.0
2     5799.0
3     1499.0
4     1999.0
...
14323   3999.0
14324   2050.0
14325   1659.0
14326   2399.0
14327   2599.0
Name: price, Length: 14328, dtype: float64
```

```
fashionset['ratingCount'] = fashionset['ratingCount'].fillna(fashionset['mc'])
```

```
fashionset
```

	p_id	Product_name	price	colour	brand_name	ratingCount	avg_rating	b
0	1518329.0	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	Dupatta Bazaar	1321.0	4.548827	

```
fashionset['avg_rating'] = fashionset['avg_rating'].fillna(fashionset['mar'])
fashionset
```

	p_id	Product_name	price	colour	brand_name	ratingCount	avg_rating	b
0	1518329.0	Dupatta Bazaar White Embroidered Chiffon Dupatta	899.0	White	Dupatta Bazaar	1321.0	4.548827	
1	5829334.0	Roadster Women Mustard Yellow Solid Hooded Swe...	1199.0	Mustard	Roadster	5462.0	4.313255	
2	10340119.0	Inddus Peach- Coloured & Beige Unstitched Dress...	5799.0	Peach	Inddus	145.0	4.068966	
3	10856380.0	SASSAFRAS Women Black Parallel Trousers	1499.0	Black	SASSAFRAS	9124.0	4.147523	

```
fashionset.to_csv(r'filled.csv')
```

```
dfp1 = pd.read_csv(r'/content/sample_data/fsh.csv')
```

```
dfp2 = pd.read_csv(r'/content/sample_data/part2.csv',engine="python")
```

```
fashiondataset = dfp1.merge(dfp2, on='p_id', how='left')
```

```
fashiondataset.to_csv(r'fdsm.csv')
```

```
import csv
import json
```

```

# Function to convert a CSV to JSON
# Takes the file paths as arguments
def make_json(csvFilePath, jsonFilePath):

    # create a dictionary
    data = {}

    # Open a csv reader called DictReader
    with open(csvFilePath, encoding='utf-8') as csvf:
        csvReader = csv.DictReader(csvf)

        # Convert each row into a dictionary
        # and add it to data
        for rows in csvReader:

            # Assuming a column named 'No' to
            # be the primary key
            key = rows['p_id']
            data[key] = rows

    # Open a json writer, and use the json.dumps()
    # function to dump data
    with open(jsonFilePath, 'w', encoding='utf-8') as jsonf:
        jsonf.write(json.dumps(data, indent=4))

# Driver Code

# Decide the two file paths according to your
# computer system
csvFilePath = r'/content/fdsm.csv'
jsonFilePath = r'fdsjson.json'

# Call the make_json function
make_json(csvFilePath, jsonFilePath)

```

[Colab paid products](#) - [Cancel contracts here](#)

---

