Open in Colab

In [2]:
```
import pandas as pd

df = pd.read_excel("/content/Attribute DataSet (1).xlsx")
print(df)
```

```
         Dress_ID    Style    Price  Rating  Size  Season   NeckLine  \
0    1.006033e+09     Sexy      Low     4.6     M  Summer     o-neck
1    1.212192e+09   Casual      Low     0.0     L  Summer     o-neck
2    1.190381e+09  vintage     High     0.0     L  Automn     o-neck
3    9.660060e+08    Brief  Average     4.6     L  Spring     o-neck
4    8.763395e+08     cute      Low     4.5     M  Summer     o-neck
..            ...      ...      ...     ...   ...     ...        ...
495  7.133920e+08   Casual      Low     4.7     M  Spring     o-neck
496  7.225651e+08     Sexy      Low     4.3  free  Summer     o-neck
497  5.328743e+08   Casual  Average     4.7     M  Summer     v-neck
498  6.554649e+08   Casual  Average     4.6     L  winter  boat-neck
499  9.199310e+08   Casual      Low     4.4  free  Summer     v-neck

    SleeveLength waiseline       Material  FabricType  Decoration  \
0      sleevless    empire            NaN     chiffon     ruffles
1          Petal   natural     microfiber         NaN     ruffles
2           full   natural       polyster         NaN         NaN
3           full   natural           silk     chiffon  embroidary
4      butterfly   natural  chiffonfabric     chiffon         bow
..           ...       ...            ...         ...         ...
495         full   natural       polyster         NaN         NaN
496         full    empire         cotton         NaN         NaN
497         full    empire         cotton         NaN        lace
498    sleevless    empire           silk  broadcloth    applique
499        short    empire         cotton    Corduroy        lace

    Pattern Type  Recommendation
0         animal             1.0
1         animal             0.0
2          print             0.0
3          print             1.0
4            dot             0.0
..           ...             ...
495        solid             1.0
496          NaN             0.0
497        solid             1.0
498        print             1.0
499        solid             0.0

[500 rows x 14 columns]
```

In [3]:
```
print(df.head())
```

```
       Dress_ID    Style    Price  Rating Size  Season NeckLine SleeveLength
\
0  1.006033e+09     Sexy      Low     4.6    M  Summer   o-neck    sleevless
1  1.212192e+09   Casual      Low     0.0    L  Summer   o-neck        Petal
2  1.190381e+09  vintage     High     0.0    L  Automn   o-neck         full
3  9.660060e+08    Brief  Average     4.6    L  Spring   o-neck         full
4  8.763395e+08     cute      Low     4.5    M  Summer   o-neck    butterfly
```

```
4  8.765995e+08    cute      Low      4.9     M  Summer    o-neck    butterfly

  waiseline       Material FabricType  Decoration Pattern Type  Recommendati
on
0   empire             NaN    chiffon     ruffles       animal
1.0
1   natural     microfiber         NaN     ruffles       animal
0.0
2   natural       polyster         NaN         NaN        print
0.0
3   natural           silk    chiffon  embroidary        print
1.0
4   natural  chiffonfabric    chiffon         bow          dot
0.0
```

In [4]:

```python
print(df.tail())
```

```
        Dress_ID   Style    Price  Rating  Size  Season   NeckLine  \
495  713391965.0  Casual      Low     4.7     M  Spring     o-neck
496  722565148.0    Sexy      Low     4.3  free  Summer     o-neck
497  532874347.0  Casual  Average     4.7     M  Summer     v-neck
498  655464934.0  Casual  Average     4.6     L  winter  boat-neck
499  919930954.0  Casual      Low     4.4  free  Summer     v-neck

    SleeveLength waiseline Material  FabricType Decoration Pattern Type  \
495         full   natural polyster         NaN        NaN        solid
496         full    empire   cotton         NaN        NaN          NaN
497         full    empire   cotton         NaN       lace        solid
498    sleevless    empire     silk  broadcloth   applique        print
499        short    empire   cotton    Corduroy       lace        solid

     Recommendation
495             1.0
496             0.0
497             1.0
498             1.0
499             0.0
```

In [5]:

```python
pd.read_csv('https://raw.githubusercontent.com/selva86/datasets/master/Smark
```

Out[5]:

	Year	Lag1	Lag2	Lag3	Lag4	Lag5	Volume	Today	Direction
0	2001	0.381	-0.192	-2.624	-1.055	5.010	1.19130	0.959	Up
1	2001	0.959	0.381	-0.192	-2.624	-1.055	1.29650	1.032	Up
2	2001	1.032	0.959	0.381	-0.192	-2.624	1.41120	-0.623	Down
3	2001	-0.623	1.032	0.959	0.381	-0.192	1.27600	0.614	Up
4	2001	0.614	-0.623	1.032	0.959	0.381	1.20570	0.213	Up
...	...	...	...	...	...	...	...	...	...
1245	2005	0.422	0.252	-0.024	-0.584	-0.285	1.88850	0.043	Up
1246	2005	0.043	0.422	0.252	-0.024	-0.584	1.28581	-0.955	Down
1247	2005	-0.955	0.043	0.422	0.252	-0.024	1.54047	0.130	Up

1248	2005	0.130	-0.955	0.043	0.422	0.252	1.42236	-0.298	Down
1249	2005	-0.298	0.130	-0.955	0.043	0.422	1.38254	-0.489	Down

1250 rows × 9 columns

In [6]:
```
df['Recommendation']
```

Out[6]:
```
0      1.0
1      0.0
2      0.0
3      1.0
4      0.0
      ...
495    1.0
496    0.0
497    1.0
498    1.0
499    0.0
Name: Recommendation, Length: 500, dtype: float64
```

In [7]:
```
df[['Rating','Recommendation','Style']]
```

Out[7]:

	Rating	Recommendation	Style
0	4.6	1.0	Sexy
1	0.0	0.0	Casual
2	0.0	0.0	vintage
3	4.6	1.0	Brief
4	4.5	0.0	cute
...	...	...	...
495	4.7	1.0	Casual
496	4.3	0.0	Sexy
497	4.7	1.0	Casual
498	4.6	1.0	Casual
499	4.4	0.0	Casual

500 rows × 3 columns

In [8]:
```
df.describe()
```

Out[8]:

	Dress_ID	Rating	Recommendation
count	5.000000e+02	500.000000	500.000000
mean	9.055417e+08	3.528600	0.420000
std	1.736190e+08	2.005364	0.494053

min	4.442820e+08	0.000000	0.000000
25%	7.673164e+08	3.700000	0.000000
50%	9.083296e+08	4.600000	0.000000
75%	1.039534e+09	4.800000	1.000000
max	1.253973e+09	5.000000	1.000000

In [9]:
```
(df['Rating']>4.0) & (df['Style']== 'Casual')
```

Out[9]:
```
0      False
1      False
2      False
3      False
4      False
       ...
495     True
496    False
497     True
498     True
499     True
Length: 500, dtype: bool
```

In [10]:
```
df.loc[0:10, ['Dress_ID', 'Style','Price','Rating']]
```

Out[10]:

	Dress_ID	Style	Price	Rating
0	1.006033e+09	Sexy	Low	4.6
1	1.212192e+09	Casual	Low	0.0
2	1.190381e+09	vintage	High	0.0
3	9.660060e+08	Brief	Average	4.6
4	8.763395e+08	cute	Low	4.5
5	1.068332e+09	bohemian	Low	0.0
6	1.220707e+09	Casual	Average	0.0
7	1.219677e+09	Novelty	Average	0.0
8	1.113094e+09	Flare	Average	0.0
9	9.852927e+08	bohemian	Low	0.0
10	1.117294e+09	party	Average	5.0

In [11]:
```
df.iloc[0:5,1:7]
```

Out[11]:

	Style	Price	Rating	Size	Season	NeckLine
0	Sexy	Low	4.6	M	Summer	o-neck
1	Casual	Low	0.0	L	Summer	o-neck
2	vintage	High	0.0	L	Automn	o-neck

3	Brief	Average	4.6	L	Spring	o-neck
4	cute	Low	4.5	M	Summer	o-neck

In [12]:
```
df.dtypes
```

Out[12]:
```
Dress_ID          float64
Style              object
Price              object
Rating            float64
Size               object
Season             object
NeckLine           object
SleeveLength       object
waiseline          object
Material           object
FabricType         object
Decoration         object
Pattern Type       object
Recommendation    float64
dtype: object
```

In [16]:
```
df2=df[df.dtypes[(df.dtypes == 'float64') | (df.dtypes == 'int64')] .index]
df2
```

Out[16]:

	Dress_ID	Rating	Recommendation
0	1.006033e+09	4.6	1.0
1	1.212192e+09	0.0	0.0
2	1.190381e+09	0.0	0.0
3	9.660060e+08	4.6	1.0
4	8.763395e+08	4.5	0.0
...	...	...	...
495	7.133920e+08	4.7	1.0
496	7.225651e+08	4.3	0.0
497	5.328743e+08	4.7	1.0
498	6.554649e+08	4.6	1.0
499	9.199310e+08	4.4	0.0

500 rows × 3 columns

In [15]:
```
data5 = {'emp_id':[101,102,103,104],
         'salary':[12,45,53,45],
         'providentfund':[12312,3432,234,535]}
df5 = pd.DataFrame(data5)
print(df5)
```

```
   emp_id  salary  providentfund
```

```
0     101      12          12312
1     102      45           3432
2     103      53            234
3     104      45            535
```

In [14]:

```
data6 = {'emp_id': [101,102,103,104],
       'mobile_no':[12321,3423,2432,2412],
       'house_no':[12312,234,2344,123]}
df6 = pd.DataFrame(data6)
print(df6)
```

```
   emp_id  mobile_no  house_no
0     101      12321     12312
1     102       3423       234
2     103       2432      2344
3     104       2412       123
```

In [18]:

```
pd.merge(df6,df5, how = 'left')
```

Out[18]:

	emp_id	mobile_no	house_no	salary	providentfund
0	101	12321	12312	12	12312
1	102	3423	234	45	3432
2	103	2432	2344	53	234
3	104	2412	123	45	535

In [17]:

```
pd.merge(df6,df5, how = 'right')
```

Out[17]:

	emp_id	mobile_no	house_no	salary	providentfund
0	101	12321	12312	12	12312
1	102	3423	234	45	3432
2	103	2432	2344	53	234
3	104	2412	123	45	535

In [19]:

```
a=pd.merge(df5,df6)
```