NAME:PRIYANSHU MANOHAR PATIL ,DIV:A,BRANCH:EXTC,ROLLNO.:55

```
pwd
```

```
    '/content'
```

```
pip install pandas
```

```
    Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/
    Requirement already satisfied: pandas in /usr/local/lib/python3.7/dist-packages (1.3
    Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.7/dist-packages
    Requirement already satisfied: numpy>=1.17.3 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dis
    Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (fr
```

```
import pandas as pd
```

```
df = pd.read_excel("Dress Sales.xlsx")
df
```

	Dress_ID	29/8/2013	31/8/2013	2013-02-09 00:00:00	2013-04-09 00:00:00	2013-06-09 00:00:00	2013-08-09 00:00:00	2013-1 00:00
0	1006032852	2114	2274	2491	2660	2727	2887	2!
1	1212192089	151	275	570	750	813	1066	1
2	1190380701	6	7	7	7	8	8	
3	966005983	1005	1128	1326	1455	1507	1621	1(
4	876339541	996	1175	1304	1396	1432	1559	1!
...	...	...	...	...	...	...	...	
495	713391965	0	0	0	560	554	544	!
496	722565148	0	0	0	875	866	861	!
497	532874347	0	0	0	734	728	726	!
498	655464934	0	0	0	254	259	261	:
499	919930954	0	0	0	538	545	558	!

500 rows × 24 columns

```
df1= pd.read_excel('C:\\Users\\content\\Downloads\\Attribute DataSet (1).xlsx')
```

```
df1.head()
```

	Dress_ID	Style	Price	Rating	Size	Season	NeckLine	SleeveLength	waisel
0	1006032852	Sexy	Low	4.6	M	Summer	o-neck	sleevless	em
1	1212192089	Casual	Low	0.0	L	Summer	o-neck	Petal	nat
2	1190380701	vintage	High	0.0	L	Automn	o-neck	full	nat
3	966005983	Brief	Average	4.6	L	Spring	o-neck	full	nat

```
df = pd.read_excel(r"Attribute DataSet.xlsx")
```

```
type(df)
```

```
pandas.core.frame.DataFrame
```

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfl
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu

```
pd.read_excel(r"C:\Users\content\Attribute dataSet.xlsx" , sheet_name="Such", header = No
```

```
pd.read_excel(r"C:\Users\content\Attribute dataSet.xlsx" , sheet_name="Such", header = No
```

		A	B	C	D	E	F
qw	**wq**	q	e	qew	q	NaN	NaN
NaN	**NaN**	NaN	NaN	NaN	NaN	qw	NaN
ewqwe	**qwq**	wewq	qwe	qwqe	qwe	wqe	qw
q2w	**qwq**	q2w	qwe	dqwe	ddqw	weqw	NaN

```
df.head()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Petal
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	full

```
df.tail()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	full
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	full
497	532874347	Casual	Average	4.7	M	2023-02-02	Summer	v-neck	full

```
df1=pd.read_csv('haberman.csv')
df1
```

	30	64	1	1.1
0	30	62	3	1
1	30	65	0	1
2	31	59	2	1
3	31	65	4	1
4	33	58	10	1

```
pd.read_csv('haberman.csv',names=['Age of patient','Patient year od operation', 'Number','
```

	Age of patient	Patient year od operation	Number	Survival status
0	30	64	1	1
1	30	62	3	1
2	30	65	0	1
3	31	59	2	1
4	31	65	4	1
...	...	...	...	...
301	75	62	1	1
302	76	67	0	1
303	77	65	3	1
304	78	65	1	2
305	83	58	2	2

306 rows × 4 columns

```
pd.read_csv('haberman1.csv')
```

	30@64@1@1
0	30@62@3@1
1	30@65@0@1

```
pd.read_csv('haberman1.csv',sep='@')
```

	30	64	1	1.1
0	30	62	3	1
1	30	65	0	1
2	31	59	2	1
3	31	65	4	1
4	33	58	10	1
...	...	...	...	...
300	75	62	1	1
301	76	67	0	1
302	77	65	3	1
303	78	65	1	2
304	83	58	2	2

305 rows × 4 columns

```
pd.read_csv('https://raw.githubusercontent.com/selva86/datasets/master/Smarket.csv')
```

```
a = pd.read_html("https://www.basketball-reference.com/leagues/NBA_2015_totals.html")
```

| **3** | 2001 | -0.623 | 1.032 | 0.959 | 0.381 | -0.192 | 1.27600 | 0.614 | | Up |

```
type(a)
```

```
list
```

| **3** | 2001 | -0.623 | 1.032 | 0.959 | 0.381 | -0.192 | 1.27600 | 0.614 | | Up |

```
len(a)
```

```
1
```

| **1245** | 2005 | 0.422 | 0.252 | 0.024 | 0.584 | 0.285 | 1.88850 | 0.043 | | Up |

```
a[0]
```

	Rk	Player	Pos	Age	Tm	G	GS	MP	FG	FGA	...	FT%	ORB	DRB	TRB
0	1	Quincy Acy	PF	24	NYK	68	22	1287	152	331	...	.784	79	222	301
1	2	Jordan Adams	SG	20	MEM	30	0	248	35	86	...	.609	9	19	28
2	3	Steven Adams	C	21	OKC	70	67	1771	217	399	...	.502	199	324	523
3	4	Jeff Adrien	PF	28	MIN	17	0	215	19	44	...	.579	23	54	77
4	5	Arron Afflalo	SG	29	TOT	78	72	2502	375	884	...	.843	27	220	247
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
670	490	Thaddeus Young	PF	26	TOT	76	68	2434	451	968	...	.655	127	284	411
671	490	Thaddeus Young	PF	26	MIN	48	48	1605	289	641	...	.682	75	170	245

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleeveless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu

df.columns

```
Index(['Dress_ID', 'Style', 'Price', 'Rating', 'Size', 'Date', 'Season',
       'NeckLine', 'SleeveLength', 'waiseline', 'Material', 'FabricType',
       'Decoration', 'Pattern Type', 'Recommendation'],
      dtype='object')
```
                                                              02-28

df['Dress_ID']

```
0      1006032852
1      1212192089
2      1190380701
3       966005983
4       876339541
          ...
495     713391965
496     722565148
497     532874347
498     655464934
499     919930954
Name: Dress_ID, Length: 500, dtype: int64
```

df1 = df['Style']

type(df)

```
pandas.core.frame.DataFrame
```

type(df1)

```
pandas.core.series.Series
```

df.dtypes

```
Dress_ID                int64
Style                  object
Price                  object
Rating                float64
Size                   object
Date           datetime64[ns]
Season                 object
NeckLine               object
SleeveLength           object
```

```
    waiseline                 object
    Material                  object
    FabricType                object
    Decoration                object
    Pattern Type              object
    Recommendation             int64
    dtype: object
```

df['Recommendation']

```
    0      1
    1      0
    2      0
    3      1
    4      0
          ..
    495    1
    496    0
    497    1
    498    1
    499    0
    Name: Recommendation, Length: 500, dtype: int64
```

df[['Rating','Recommendation','Style']]

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

500 rows × 3 columns

df.describe()

	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

df.dtypes

```
Dress_ID                    int64
Style                      object
Price                      object
Rating                    float64
Size                       object
Date               datetime64[ns]
Season                     object
NeckLine                   object
SleeveLength               object
waiseline                  object
Material                   object
FabricType                 object
Decoration                 object
Pattern Type               object
Recommendation              int64
dtype: object
```

df.dtypes == 'object'

```
Dress_ID          False
Style              True
Price              True
Rating            False
Size               True
Date              False
Season             True
NeckLine           True
SleeveLength       True
waiseline          True
Material           True
FabricType         True
Decoration         True
Pattern Type       True
Recommendation    False
dtype: bool
```

df.dtypes[df.dtypes == 'object']

```
Style          object
Price          object
Size           object
Season         object
NeckLine       object
```

```
       SleeveLength    object
       waiseline       object
       Material        object
       FabricType      object
       Decoration      object
       Pattern Type    object
       dtype: object
```

```
df.dtypes[df.dtypes == 'object'].index
```

```
Index(['Style', 'Price', 'Size', 'Season', 'NeckLine', 'SleeveLength',
       'waiseline', 'Material', 'FabricType', 'Decoration', 'Pattern Type'],
      dtype='object')
```

```
df[df.dtypes[df.dtypes == 'object'].index]
```

	Style	Price	Size	Season	NeckLine	SleeveLength	waiseline	Material	Fa
0	Sexy	Low	M	Summer	o-neck	sleevless	empire	NaN	
1	Casual	Low	L	Summer	o-neck	Petal	natural	microfiber	
2	vintage	High	L	Automn	o-neck	full	natural	polyster	
3	Brief	Average	L	Spring	o-neck	full	natural	silk	
4	cute	Low	M	Summer	o-neck	butterfly	natural	chiffonfabric	
...	...	...	...	...	...	...	...	...	
495	Casual	Low	M	Spring	o-neck	full	natural	polyster	
496	Sexy	Low	free	Summer	o-neck	full	empire	cotton	
497	Casual	Average	M	Summer	v-neck	full	empire	cotton	
498	Casual	Average	L	winter	boat-neck	sleevless	empire	silk	
499	Casual	Low	free	Summer	v-neck	short	empire	cotton	

```
df[df.dtypes[df.dtypes == 'object'].index].describe()
```

	Style	Price	Size	Season	NeckLine	SleeveLength	waiseline	Material	
count	500	498	500	498	497	498	413	372	
unique	13	7	7	8	16	17	4	23	
top	Casual	Average	M	Summer	o-neck	sleevless	natural	cotton	

```
df.dtypes[df.dtypes == 'float']
```

```
       Rating    float64
```

```
        dtype: object
```

```
df[df.dtypes[df.dtypes == 'float'].index]
```

	Rating
0	4.6
1	0.0
2	0.0
3	4.6
4	4.5
...	...
495	4.7
496	4.3
497	4.7
498	4.6
499	4.4

500 rows × 1 columns

```
df[df.dtypes[df.dtypes == 'float'].index].describe
```

```
<bound method NDFrame.describe of      Rating
0      4.6
1      0.0
2      0.0
3      4.6
4      4.5
..     ...
495    4.7
496    4.3
497    4.7
498    4.6
499    4.4

[500 rows x 1 columns]>
```

```
df['Dress_ID']
```

```
0      1006032852
1      1212192089
2      1190380701
3       966005983
4       876339541
          ...
495     713391965
496     722565148
497     532874347
498     655464934
```

```
499    919930954
Name: Dress_ID, Length: 500, dtype: int64
```

```
df['Dress_ID'] [1:14:2]
```

```
1     1212192089
3      966005983
5     1068332458
7     1219677488
9      985292672
11     898481530
13     749031896
Name: Dress_ID, dtype: int64
```

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevles
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevles
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	shor

500 rows × 15 columns

```
df['category']='Suchi'
```

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLengt
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevles:
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfl
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevles:
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	shor

```
df.columns
```

```
Index(['Dress_ID', 'Style', 'Price', 'Rating', 'Size', 'Date', 'Season',
       'NeckLine', 'SleeveLength', 'waiseline', 'Material', 'FabricType',
       'Decoration', 'Pattern Type', 'Recommendation', 'category'],
      dtype='object')
```

```
df['Recommendation'].isnull()
```

```
0      False
1      False
2      False
3      False
4      False
       ...
495    False
496    False
497    False
498    False
499    False
Name: Recommendation, Length: 500, dtype: bool
```

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevless
						2023-			

```
df.columns
```

```
Index(['Dress_ID', 'Style', 'Price', 'Rating', 'Size', 'Date', 'Season',
       'NeckLine', 'SleeveLength', 'waiseline', 'Material', 'FabricType',
       'Decoration', 'Pattern Type', 'Recommendation', 'category'],
      dtype='object')
```

```
df['Rating']== max(df['Rating'])
```

```
0      False
1      False
2      False
3      False
4      False
       ...
495    False
496    False
497    False
498    False
499    False
Name: Rating, Length: 500, dtype: bool
```

```
df['Material']
```

```
0             NaN
1       microfiber
2         polyster
3             silk
4      chiffonfabric
         ...
495       polyster
```

```
496          cotton
497          cotton
498           silk
499          cotton
Name: Material, Length: 500, dtype: object
```

```
df[df['Material']== 'cotton']
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLen
6	1220707172	Casual	Average	0.0	XL	NaT	Summer	o-neck	
8	1113094204	Flare	Average	0.0	free	NaT	Spring	v-neck	sl
13	749031896	vintage	Average	4.8	M	NaT	Summer	o-neck	sl
14	1055411544	Casual	Low	5.0	M	NaT	Summer	boat-neck	sl
16	624314841	cute	Average	4.7	L	NaT	spring	o-neck	sl
...	...	...	...	...	...	...	...	...	
491	964917582	Casual	Average	5.0	L	2023-02-24	Summer	o-neck	sleevl
493	817353671	bohemian	Low	4.6	free	2023-02-26	Summer	o-neck	sleevl
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	sl

152 rows × 16 columns

```
len(df[df['Material']== 'cotton'])
```

```
152
```

```
df[df['Rating']>4.5]
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLeng
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevl
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	
10	1117293701	party	Average	5.0	free	NaT	Summer	o-neck	
12	957723897	sexy	Low	4.7	M	NaT	Winter	o-neck	threequa
13	749031896	vintage	Average	4.8	M	NaT	Summer	o-neck	sl
...	...	...	...	...	...	...	...	...	
493	817353671	bohemian	Low	4.6	free	2023-02-26	Summer	o-neck	sleevl
494	990559192	Brief	Average	4.7	M	2023-02-27	winter	o-neck	halfsle

```
df[df['Rating']>4.5]['Style']
```

```
0         Sexy
3        Brief
10       party
12        sexy
13      vintage
         ...
493    bohemian
494       Brief
495      Casual
497      Casual
498      Casual
Name: Style, Length: 275, dtype: object
```

```
(df['Rating']>4.5) & (df['Style']== 'Sexy')
```

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

```
df.dtypes
```

```
Dress_ID             int64
Style               object
Price               object
Rating             float64
Size                object
```
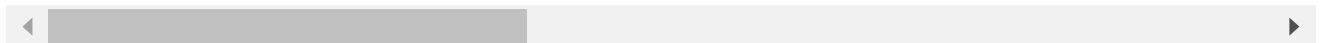
```
Date              datetime64[ns]
Season                    object
NeckLine                  object
SleeveLength              object
waiseline                 object
Material                  object
FabricType                object
Decoration                object
Pattern Type              object
Recommendation             int64
category                  object
dtype: object
```

```python
pd.read_excel("Dress Sales.xlsx")
```

	Dress_ID	29/8/2013	31/8/2013	2013-02-09 00:00:00	2013-04-09 00:00:00	2013-06-09 00:00:00	2013-08-09 00:00:00	2013-1 00:00
0	1006032852	2114	2274	2491	2660	2727	2887	2!
1	1212192089	151	275	570	750	813	1066	1
2	1190380701	6	7	7	7	8	8	
3	966005983	1005	1128	1326	1455	1507	1621	1(
4	876339541	996	1175	1304	1396	1432	1559	1!
...	...	...	...	...	...	...	...	
495	713391965	0	0	0	560	554	544	!
496	722565148	0	0	0	875	866	861	{
497	532874347	0	0	0	734	728	726	`
498	655464934	0	0	0	254	259	261	:
499	919930954	0	0	0	538	545	558	!

500 rows × 24 columns

```python
df = pd.read_excel("Attribute DataSet.xlsx")
```

```python
date = pd.to_datetime(df['Date'])
```
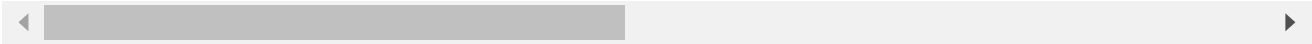
```python
type(date[0])
```

```
pandas._libs.tslibs.nattype.NaTType
```

```python
df['converted_order_date'] = pd.to_datetime(df['Date'])
```

```python
df
```

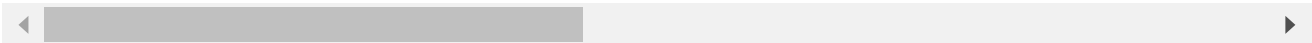|	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevless
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	shor

500 rows × 16 columns

```
df['order_date_year'] = df['converted_order_date'].dt.year
```

```
df.head()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Petal
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	full
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly

```
df.tail()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	full
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	full
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	full
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevless
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	short

```
df['order_date_month'] = df['converted_order_date'].dt.month
```

```
df.tail()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	full
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	full
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	full
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevless
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	short

```
df['order_date_week'] = df['converted_order_date'].dt.week
```

```
C:\Users\SUCHIT~1\AppData\Local\Temp/ipykernel_14724/4037216944.py:1: FutureWarning:
  df['order_date_week'] = df['converted_order_date'].dt.week
```
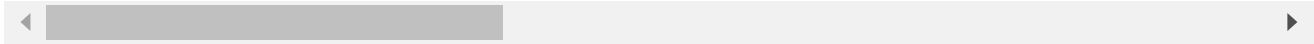
```
df.tail()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	full
496	722565148	Sexy	Low	4.3	free	2023-	Summer	o-neck	full

```
df.head()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Petal
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	full
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly

```
df[['converted_order_date', 'order_date_week']]
```

	converted_order_date	order_date_week
0	NaT	NaN
1	NaT	NaN
2	NaT	NaN
3	NaT	NaN
4	NaT	NaN
...	...	...
495	2023-02-28	9.0
496	2023-03-01	9.0
497	2023-03-02	9.0
498	2023-03-03	9.0
499	2023-03-04	9.0

500 rows × 2 columns

```
df['order_date_year']==2023
```

```
0     False
1     False
2     False
3     False
4     False
```

```
           ...
495        True
496        True
497        True
498        True
499        True
Name: order_date_year, Length: 500, dtype: bool
```
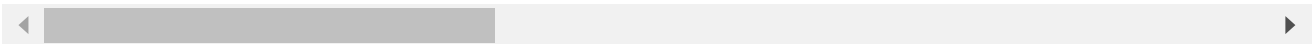
```
df['order_date_month'].value_counts()
```

```
1.0      62
2.0      56
12.0     47
3.0      35
5.0      31
7.0      31
8.0      31
10.0     31
4.0      30
6.0      30
9.0      30
11.0     30
Name: order_date_month, dtype: int64
```

```
df
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Peta
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	fu
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	fu
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly
...	...	...	...	...	...	...	...	...	..
495	713391965	Casual	Low	4.7	M	2023-02-28	Spring	o-neck	fu
496	722565148	Sexy	Low	4.3	free	2023-03-01	Summer	o-neck	fu
497	532874347	Casual	Average	4.7	M	2023-03-02	Summer	v-neck	fu
498	655464934	Casual	Average	4.6	L	2023-03-03	winter	boat-neck	sleevless
499	919930954	Casual	Low	4.4	free	2023-03-04	Summer	v-neck	shor

500 rows × 19 columns

```
df.head()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Petal
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	full
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly

```
df.head()
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
0	1006032852	Sexy	Low	4.6	M	NaT	Summer	o-neck	sleevless
1	1212192089	Casual	Low	0.0	L	NaT	Summer	o-neck	Petal
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
3	966005983	Brief	Average	4.6	L	NaT	Spring	o-neck	full
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly

```
df.loc[2:10:2]
```

	Dress_ID	Style	Price	Rating	Size	Date	Season	NeckLine	SleeveLength
2	1190380701	vintage	High	0.0	L	NaT	Automn	o-neck	full
4	876339541	cute	Low	4.5	M	NaT	Summer	o-neck	butterfly
6	1220707172	Casual	Average	0.0	XL	NaT	Summer	o-neck	full
8	1113094204	Flare	Average	0.0	free	NaT	Spring	v-neck	short
10	1117293701	party	Average	5.0	free	NaT	Summer	o-neck	full

```
df.loc[0:3, ['Dress_ID', 'Style','Price']]
```

	Dress_ID	Style	Price
0	1006032852	Sexy	Low

```
df.iloc[0:5,4:7]
```

	Size	Date	Season
0	M	NaT	Summer
1	L	NaT	Summer
2	L	NaT	Automn
3	L	NaT	Spring
4	M	NaT	Summer

```
#loc=named indexes, iloc default indexes
```

```
df.iloc[2:6,4:7]
```

	Size	Date	Season
2	L	NaT	Automn
3	L	NaT	Spring
4	M	NaT	Summer
5	M	NaT	Summer

```
df.dtypes
```

```
Dress_ID                      int64
Style                        object
Price                        object
Rating                      float64
Size                         object
Date                 datetime64[ns]
Season                       object
NeckLine                     object
SleeveLength                 object
waiseline                    object
Material                     object
FabricType                   object
Decoration                   object
Pattern Type                 object
Recommendation                int64
converted_order_date datetime64[ns]
order_date_year             float64
order_date_month            float64
order_date_week             float64
dtype: object
```

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

df2

	Dress_ID	Rating	Recommendation
0	1006032852	4.6	1
1	1212192089	0.0	0
2	1190380701	0.0	0
3	966005983	4.6	1
4	876339541	4.5	0
...	...	...	...
495	713391965	4.7	1
496	722565148	4.3	0
497	532874347	4.7	1
498	655464934	4.6	1
499	919930954	4.4	0

500 rows × 3 columns

```
df3 = df2[df2 == 1]
```

```
df3.dropna(axis = 1)
```

0
1
2
3
4
...
495
496
497
498
499

500 rows × 0 columns

```
df3.dropna(thresh=1)
```

	Dress_ID	Rating	Recommendation
0	NaN	NaN	1.0
3	NaN	NaN	1.0
8	NaN	NaN	1.0
9	NaN	NaN	1.0
12	NaN	NaN	1.0
...	...	...	...
491	NaN	NaN	1.0
493	NaN	NaN	1.0
495	NaN	NaN	1.0
497	NaN	NaN	1.0
498	NaN	NaN	1.0

211 rows × 3 columns

```
df3
```

	Dress_ID	Rating	Recommendation
0	NaN	NaN	1.0
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	1.0
4	NaN	NaN	NaN
...	...	...	...
495	NaN	NaN	1.0
496	NaN	NaN	NaN
497	NaN	NaN	1.0
498	NaN	NaN	1.0
499	NaN	NaN	NaN

500 rows × 3 columns

```
df3.fillna(value=3)
```

	Dress_ID	Rating	Recommendation
0	3.0	3.0	1.0
1	3.0	3.0	3.0
2	3.0	3.0	3.0
3	3.0	3.0	1.0
4	3.0	3.0	3.0
...	...	...	...
495	3.0	3.0	1.0
496	3.0	3.0	3.0
497	3.0	3.0	1.0
498	3.0	3.0	1.0

```
df3.fillna(value=df3.mean['Reccommendation'])
```

	Dress_ID	Rating	Recommendation
0	NaN	NaN	1.0
1	NaN	NaN	NaN
2	NaN	NaN	NaN
3	NaN	NaN	1.0
4	NaN	NaN	NaN
...	...	...	...
495	NaN	NaN	1.0
496	NaN	NaN	NaN
497	NaN	NaN	1.0
498	NaN	NaN	1.0
499	NaN	NaN	NaN

500 rows × 3 columns

```
df2
```

	Dress_ID	Rating	Recommendation
0	1006032852	4.6	1
1	1212192089	0.0	0
2	1190380701	0.0	0
3	966005983	4.6	1
4	876339541	4.5	0
...	...	...	...
495	713391965	4.7	1

```
df3 = df2.groupby('Rating')
```

```
df3
```

```
<pandas.core.groupby.generic.DataFrameGroupBy object at 0x000001D8A32C1400>
```

```
df2.columns
```

```
Index(['Dress_ID', 'Rating', 'Recommendation'], dtype='object')
```

```
df.dtypes
```

```
Dress_ID               int64
Style                 object
Price                 object
Rating               float64
Size                  object
Date          datetime64[ns]
Season                object
NeckLine              object
SleeveLength          object
waiseline             object
Material              object
FabricType            object
Decoration            object
Pattern Type          object
Recommendation         int64
dtype: object
```

```
data = {'name' : ['Sudh', 'Krish','Nitesh','Tulesco'],
        'salary': [100, 200, 300, 400],
        'email': ['spatil@mes.ac.in', 'krish@mes.ac.in', 'nitesh@mes.ac.in','tilesco@mes.ac
        'addr': ['wer', 'ert', 'ert', 'weew']}
```

```
df = pd.DataFrame(data)
```

```
pd.DataFrame(data,index = [4,5,6,7])
```

	name	salary	email	addr
4	Sudh	100	spatil@mes.ac.in	wer
5	Krish	200	krish@mes.ac.in	ert
6	Nitesh	300	nitesh@mes.ac.in	ert
7	Tulesco	400	tilesco@mes.ac.in	weew

```
df.loc[5:6]
```

	name	salary	email	addr

```
df.iloc[1:3]
```

	name	salary	email	addr
1	Krish	200	krish@mes.ac.in	ert
2	Nitesh	300	nitesh@mes.ac.in	ert

```
pd.DataFrame(data,index = [0,1,2,3])
```

	name	salary	email	addr
0	Sudh	100	spatil@mes.ac.in	wer
1	Krish	200	krish@mes.ac.in	ert
2	Nitesh	300	nitesh@mes.ac.in	ert
3	Tulesco	400	tilesco@mes.ac.in	weew

```
data1 = {'pf_num':[12,23,34,54],
        'incometax': [132,435,452,123],
         'mobile': [1324123,2314123,341,23423],
         'courses':['ds','big data', 'dl','python']}
```

```
df1 = pd.DataFrame(data1)
```

```
df1
```

```
df1
```

	pf_num	incometax	mobile	courses

```
pd.concat([df,df1])
```

	name	salary	email	addr	pf_num	incometax	mobile	courses
0	Sudh	100.0	spatil@mes.ac.in	wer	NaN	NaN	NaN	NaN
1	Krish	200.0	krish@mes.ac.in	ert	NaN	NaN	NaN	NaN
2	Nitesh	300.0	nitesh@mes.ac.in	ert	NaN	NaN	NaN	NaN
3	Tulesco	400.0	tilesco@mes.ac.in	weew	NaN	NaN	NaN	NaN
0	NaN	NaN	NaN	NaN	12.0	132.0	1324123.0	ds
1	NaN	NaN	NaN	NaN	23.0	435.0	2314123.0	big data
2	NaN	NaN	NaN	NaN	34.0	452.0	341.0	dl
3	NaN	NaN	NaN	NaN	54.0	123.0	23423.0	python

```
pd.concat([df,df1], axis=1)
```

	name	salary	email	addr	pf_num	incometax	mobile	courses
0	Sudh	100	spatil@mes.ac.in	wer	12	132	1324123	ds
1	Krish	200	krish@mes.ac.in	ert	23	435	2314123	big data
2	Nitesh	300	nitesh@mes.ac.in	ert	34	452	341	dl
3	Tulesco	400	tilesco@mes.ac.in	weew	54	123	23423	python

```
pd.concat([df,df1], axis=0)
```

	name	salary	email	addr	pf_num	incometax	mobile	courses
0	Sudh	100.0	spatil@mes.ac.in	wer	NaN	NaN	NaN	NaN
1	Krish	200.0	krish@mes.ac.in	ert	NaN	NaN	NaN	NaN
2	Nitesh	300.0	nitesh@mes.ac.in	ert	NaN	NaN	NaN	NaN
3	Tulesco	400.0	tilesco@mes.ac.in	weew	NaN	NaN	NaN	NaN
0	NaN	NaN	NaN	NaN	12.0	132.0	1324123.0	ds
1	NaN	NaN	NaN	NaN	23.0	435.0	2314123.0	big data
2	NaN	NaN	NaN	NaN	34.0	452.0	341.0	dl
3	NaN	NaN	NaN	NaN	54.0	123.0	23423.0	python

```
data2 = {'0' : ['Sudh', 'Krish','Nitesh','Tulesco'],
      '1': [100, 200, 300, 400],
```

```
        '2': ['spatil@mes.ac.in', 'krish@mes.ac.in', 'nitesh@mes.ac.in','tilesco@mes.ac.in'
        '3': ['wer', 'ert', 'ert', 'weew']}
data3 = {'0':[12,23,34,54],
        '1': [132,435,452,123],
         '2': [1324123,2314123,341,23423],
         '3':['ds','big data', 'dl','python']}
```

```
df3=pd.DataFrame(data2)
```

```
df4=pd.DataFrame(data3)
```

```
pd.concat([df3,df4])
```

	0	1	2	3
0	Sudh	100	spatil@mes.ac.in	wer
1	Krish	200	krish@mes.ac.in	ert
2	Nitesh	300	nitesh@mes.ac.in	ert
3	Tulesco	400	tilesco@mes.ac.in	weew
0	12	132	1324123	ds
1	23	435	2314123	big data
2	34	452	341	dl
3	54	123	23423	python

```
data5 = {'emp_id':[101,102,103,104],
        'salary':[12,45,53,45],
        'providentfund':[12312,3432,234,535]}
```

```
data6 = {'emp_id': [101,102,103,104],
        'mobile_no':[12321,3423,2432,2412],
        'house_no':[12312,234,2344,123]}
```

```
df5 = pd.DataFrame(data5)
df5
```

	emp_id	salary	providentfund
0	101	12	12312
1	102	45	3432
2	103	53	234
3	104	45	535

```
df6 = pd.DataFrame(data6)
```

```
df6
```

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

```
df6
```

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

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

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

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

	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

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

	emp_id	mobile_no	house_no	salary	providentfund
0	101	12321	12312	12	12312
1	102	3423	234	45	3432

```
data7 = {'emp_id1':[101,102,103,104],
         'salary':[12,45,53,45],
         'providentfund':[12312,3432,234,535]}
data8 = {'emp_id2': [101,102,103,104],
         'mobile_no':[12321,3423,2432,2412],
         'house_no':[12312,234,2344,123]}
```

```
df7 = pd.DataFrame(data7)
df8 = pd.DataFrame(data8)
```

```
df7
```

	emp_id1	salary	providentfund
0	101	12	12312
1	102	45	3432
2	103	53	234
3	104	45	535

```
pd.merge(df7,df8, left_on='emp_id1', right_on='emp_id2', how = 'inner')
```

	emp_id1	salary	providentfund	emp_id2	mobile_no	house_no
0	101	12	12312	101	12321	12312
1	102	45	3432	102	3423	234
2	103	53	234	103	2432	2344
3	104	45	535	104	2412	123

```
data9 = {'emp_id':[101,102,103,104],
         'salary':[12,45,53,45],
         'providentfund':[12312,3432,234,535]}
data10 = {'emp_id': [101,102,104,103],
          'mobile_no':[12321,3423,2432,2412],
          'house_no':[12312,234,2344,123]}
```

```
df9 = pd.DataFrame(data9)
df10 = pd.DataFrame(data10)
```

```
df = pd.merge(df9,df10, on = ['emp_id'])
```

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

```
df = pd.read_csv("pollution.csv")
```

```
df[2000:3000]
```

	No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir
2000	2001	2010	3	25	8	12.0	-16	3.0	1030.0	NW	21.01	0	0
2001	2002	2010	3	25	9	10.0	-15	4.0	1030.0	NW	28.16	0	0
2002	2003	2010	3	25	10	9.0	-17	5.0	1030.0	NW	37.99	0	0
2003	2004	2010	3	25	11	12.0	-20	6.0	1029.0	NW	46.93	0	0
2004	2005	2010	3	25	12	12.0	-20	7.0	1028.0	NW	51.85	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
2995	2996	2010	5	5	19	14.0	0	18.0	1000.0	NW	84.04	0	0
2996	2997	2010	5	5	20	14.0	0	17.0	1002.0	NW	97.00	0	0
2997	2998	2010	5	5	21	14.0	-1	17.0	1003.0	NW	113.09	0	0
2998	2999	2010	5	5	22	12.0	-1	16.0	1004.0	NW	124.27	0	0
2999	3000	2010	5	5	23	10.0	-1	15.0	1004.0	NW	134.10	0	0

1000 rows × 13 columns

```
def profit_flag(a):
    if a>0:
        return 'positive'
    else :
        return 'negative'
```

```
profit_flag(-23)
```

```
    'negative'
```

```
profit_flag(23)
```

```
    'positive'
```

```
df['TEMP_new']=df['TEMP'].apply(profit_flag)
```

```
df
```

	No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir
0	1	2010	1	1	0	NaN	-21	-11.0	1021.0	NW	1.79	0	0
1	2	2010	1	1	1	NaN	-21	-12.0	1020.0	NW	4.92	0	0
2	3	2010	1	1	2	NaN	-21	-11.0	1019.0	NW	6.71	0	0
3	4	2010	1	1	3	NaN	-21	-14.0	1019.0	NW	9.84	0	0
4	5	2010	1	1	4	NaN	-20	-12.0	1018.0	NW	12.97	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
43819	43820	2014	12	31	19	8.0	-23	-2.0	1034.0	NW	231.97	0	0
43820	43821	2014	12	31	20	10.0	-22	-3.0	1034.0	NW	237.78	0	0
43821	43822	2014	12	31	21	10.0	-22	-3.0	1034.0	NW	242.70	0	0
43822	43823	2014	12	31	22	8.0	-22	-4.0	1034.0	NW	246.72	0	0
43823	43824	2014	12	31	23	12.0	-21	-3.0	1034.0	NW	249.85	0	0

43824 rows × 14 columns

```
df.head()
```

	No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir	TEMP_ne
0	1	2010	1	1	0	NaN	-21	-11.0	1021.0	NW	1.79	0	0	negativ
1	2	2010	1	1	1	NaN	-21	-12.0	1020.0	NW	4.92	0	0	negativ
2	3	2010	1	1	2	NaN	-21	-11.0	1019.0	NW	6.71	0	0	negativ
3	4	2010	1	1	3	NaN	-21	-14.0	1019.0	NW	9.84	0	0	negativ
4	5	2010	1	1	4	NaN	-20	-12.0	1018.0	NW	12.97	0	0	negativ

```
def hour_flag(a):
    if a>10:
        return 'low'
    elif a>10 and a<20:
        return 'medium'
    else:
        return 'high'
```

```
df['flag_Hour']=df['hour'].apply(hour_flag)
```

```
df.head(50)
```

	No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir	TEMP_n
0	1	2010	1	1	0	NaN	-21	-11.0	1021.0	NW	1.79	0	0	negat
1	2	2010	1	1	1	NaN	-21	-12.0	1020.0	NW	4.92	0	0	negat
2	3	2010	1	1	2	NaN	-21	-11.0	1019.0	NW	6.71	0	0	negat
3	4	2010	1	1	3	NaN	-21	-14.0	1019.0	NW	9.84	0	0	negat
4	5	2010	1	1	4	NaN	-20	-12.0	1018.0	NW	12.97	0	0	negat
5	6	2010	1	1	5	NaN	-19	-10.0	1017.0	NW	16.10	0	0	negat
6	7	2010	1	1	6	NaN	-19	-9.0	1017.0	NW	19.23	0	0	negat
7	8	2010	1	1	7	NaN	-19	-9.0	1017.0	NW	21.02	0	0	negat
8	9	2010	1	1	8	NaN	-19	-9.0	1017.0	NW	24.15	0	0	negat
9	10	2010	1	1	9	NaN	-20	-8.0	1017.0	NW	27.28	0	0	negat
10	11	2010	1	1	10	NaN	-19	-7.0	1017.0	NW	31.30	0	0	negat
11	12	2010	1	1	11	NaN	-18	-5.0	1017.0	NW	34.43	0	0	negat
12	13	2010	1	1	12	NaN	-19	-5.0	1015.0	NW	37.56	0	0	negat
13	14	2010	1	1	13	NaN	-18	-3.0	1015.0	NW	40.69	0	0	negat
14	15	2010	1	1	14	NaN	-18	-2.0	1014.0	NW	43.82	0	0	negat
15	16	2010	1	1	15	NaN	-18	-1.0	1014.0	cv	0.89	0	0	negat
16	17	2010	1	1	16	NaN	-19	-2.0	1015.0	NW	1.79	0	0	negat
17	18	2010	1	1	17	NaN	-18	-3.0	1015.0	NW	2.68	0	0	negat

```
#derived column
df['square_DEWP']=df['DEWP'].apply(lambda a:a**2)


df
```