

## exercise3

November 2, 2025

```
[138]: import pandas as pd  
df=pd.read_csv('pre_process_datasample.csv')
```

```
[139]: df
```

```
[139]:   Country    Age    Salary Purchased  
0     France  44.0  72000.0      No  
1     Spain   27.0  48000.0     Yes  
2   Germany  30.0  54000.0      No  
3     Spain   38.0  61000.0      No  
4   Germany  40.0  61000.0     Yes  
5     France  35.0  58000.0     Yes  
6     Spain   44.0  52000.0      No  
7     France  48.0  79000.0     Yes  
8     France  50.0  83000.0      No  
9     France  37.0  67000.0     Yes
```

```
[140]: df['Country'].isnull()
```

```
[140]: 0    False  
1    False  
2    False  
3    False  
4    False  
5    False  
6    False  
7    False  
8    False  
9    False  
Name: Country, dtype: bool
```

```
[141]: df.loc[8]
```

```
[141]: Country        France  
Age            50.0  
Salary        83000.0  
Purchased      No  
Name: 8, dtype: object
```

```
[142]: n=df['Country'].mode()[0]
print(n)
```

France

```
[143]: df['Country']=df['Country'].fillna(n)
```

```
[144]: df
```

```
[144]:   Country    Age    Salary Purchased
0    France  44.0  72000.0      No
1    Spain   27.0  48000.0     Yes
2  Germany  30.0  54000.0      No
3    Spain   38.0  61000.0      No
4  Germany  40.0  61000.0     Yes
5    France  35.0  58000.0     Yes
6    Spain   44.0  52000.0      No
7    France  48.0  79000.0     Yes
8    France  50.0  83000.0      No
9    France  37.0  67000.0     Yes
```

```
[145]: df['Age'].count()
```

```
[145]: np.int64(10)
```

```
[146]: df['Age'].isna().sum()
```

```
[146]: np.int64(0)
```

```
[147]: df['Age']=df['Age'].fillna(df['Age'].mode())
```

```
[148]: df
```

```
[148]:   Country    Age    Salary Purchased
0    France  44.0  72000.0      No
1    Spain   27.0  48000.0     Yes
2  Germany  30.0  54000.0      No
3    Spain   38.0  61000.0      No
4  Germany  40.0  61000.0     Yes
5    France  35.0  58000.0     Yes
6    Spain   44.0  52000.0      No
7    France  48.0  79000.0     Yes
8    France  50.0  83000.0      No
9    France  37.0  67000.0     Yes
```

```
[149]: df['Salary']=df['Salary'].fillna(df['Salary'].mode())
```

```
[150]: df
```

```
[150]:   Country    Age    Salary Purchased
      0   France    44.0   72000.0      No
      1   Spain     27.0   48000.0     Yes
      2  Germany    30.0   54000.0      No
      3   Spain     38.0   61000.0      No
      4  Germany    40.0   61000.0     Yes
      5   France    35.0   58000.0     Yes
      6   Spain     44.0   52000.0      No
      7   France    48.0   79000.0     Yes
      8   France    50.0   83000.0      No
      9   France    37.0   67000.0     Yes
```

```
[153]: df
```

```
[153]:   Country    Age    Salary Purchased
      0   France    44.0   72000.0      No
      1   Spain     27.0   48000.0     Yes
      2  Germany    30.0   54000.0      No
      3   Spain     38.0   61000.0      No
      4  Germany    40.0   61000.0     Yes
      5   France    35.0   58000.0     Yes
      6   Spain     44.0   52000.0      No
      7   France    48.0   79000.0     Yes
      8   France    50.0   83000.0      No
      9   France    37.0   67000.0     Yes
```

```
[154]: df.to_csv('pre_process_datasample.csv',index=False)
```

```
[155]: df
```

```
[155]:   Country    Age    Salary Purchased
      0   France    44.0   72000.0      No
      1   Spain     27.0   48000.0     Yes
      2  Germany    30.0   54000.0      No
      3   Spain     38.0   61000.0      No
      4  Germany    40.0   61000.0     Yes
      5   France    35.0   58000.0     Yes
      6   Spain     44.0   52000.0      No
      7   France    48.0   79000.0     Yes
      8   France    50.0   83000.0      No
      9   France    37.0   67000.0     Yes
```

```
[ ]: #next file pre-processing
```

```
[217]: df1=pd.read_csv('Hotel_Dataset.csv')
```

```
[218]: df1
```

```
[218]:   CustomerID Age_Group Rating(1-5) Hotel FoodPreference Bill \
0           1    20-25          4     Ibis      veg  1300
1           2    30-35          5 LemonTree Non-Veg 2000
2           3    25-30          6  RedFox      Veg 1322
3           4    20-25         -1 LemonTree      Veg 1234
4           5    35+            3     Ibis Vegetarian  989
5           6    35+            3    Ibys Non-Veg 1909
6           7    35+            4  RedFox Vegetarian 1000
7           8    20-25          7 LemonTree      Veg 2999
8           9    25-30          2     Ibis Non-Veg 3456
9           9    25-30          2     Ibis Non-Veg 3456
10          10   30-35          5  RedFox non-Veg -6755
```

	NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000	20-25
1	3	59000	30-35
2	2	30000	25-30
3	2	120000	20-25
4	2	45000	35+
5	2	122220	35+
6	-1	21122	35+
7	-10	345673	20-25
8	3	-99999	25-30
9	3	-99999	25-30
10	4	87777	30-35

```
[219]: df1.drop_duplicates(inplace=True)
```

```
[220]: df1
```

```
[220]:   CustomerID Age_Group Rating(1-5) Hotel FoodPreference Bill \
0           1    20-25          4     Ibis      veg  1300
1           2    30-35          5 LemonTree Non-Veg 2000
2           3    25-30          6  RedFox      Veg 1322
3           4    20-25         -1 LemonTree      Veg 1234
4           5    35+            3     Ibis Vegetarian  989
5           6    35+            3    Ibys Non-Veg 1909
6           7    35+            4  RedFox Vegetarian 1000
7           8    20-25          7 LemonTree      Veg 2999
8           9    25-30          2     Ibis Non-Veg 3456
10          10   30-35          5  RedFox non-Veg -6755
```

	NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000	20-25
1	3	59000	30-35
2	2	30000	25-30
3	2	120000	20-25

```

4          2        45000      35+
5          2       122220      35+
6         -1        21122      35+
7         -10       345673    20-25
8          3       -99999     25-30
10         4        87777     30-35

```

[221]: df1.reset\_index(drop=True, inplace=False)

```

[221]:   CustomerID  Age_Group  Rating(1-5)      Hotel FoodPreference  Bill  NoOfPax \
0           1    20-25          4      Ibis      veg  1300        2
1           2    30-35          5  LemonTree  Non-Veg  2000        3
2           3    25-30          6     RedFox      Veg  1322        2
3           4    20-25         -1  LemonTree      Veg  1234        2
4           5      35+          3      Ibis  Vegetarian  989        2
5           6      35+          3      Ibys  Non-Veg  1909        2
6           7      35+          4     RedFox  Vegetarian 1000       -1
7           8    20-25          7  LemonTree      Veg  2999       -10
8           9    25-30          2      Ibis  Non-Veg  3456        3
9          10    30-35          5     RedFox  non-Veg -6755        4

```

```

EstimatedSalary  Age_Group.1
0        40000    20-25
1        59000    30-35
2        30000    25-30
3       120000    20-25
4        45000      35+
5       122220      35+
6        21122      35+
7       345673    20-25
8       -99999     25-30
9        87777    30-35

```

[222]: df1['FoodPreference']=df1['FoodPreference'].str.replace('Vegetarian', 'Veg')

[223]: df1['FoodPreference']=df1['FoodPreference'].str.replace('non-veg','Non-Veg')

[224]: df1

```

[224]:   CustomerID  Age_Group  Rating(1-5)      Hotel FoodPreference  Bill  \
0           1    20-25          4      Ibis      veg  1300
1           2    30-35          5  LemonTree  Non-Veg  2000
2           3    25-30          6     RedFox      Veg  1322
3           4    20-25         -1  LemonTree      Veg  1234
4           5      35+          3      Ibis      Veg  989
5           6      35+          3      Ibys  Non-Veg  1909
6           7      35+          4     RedFox      Veg  1000

```

7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	non-Veg	-6755

NoOfPax	EstimatedSalary	Age_Group.1
0	40000	20-25
1	59000	30-35
2	30000	25-30
3	120000	20-25
4	45000	35+
5	122220	35+
6	21122	35+
7	345673	20-25
8	-99999	25-30
10	87777	30-35

```
[225]: df1['FoodPreference']=df1['FoodPreference'].str.replace('veg', 'Veg')
```

```
[226]: df1
```

CustomerID	Age_Group	Rating(1-5)	Hotel	FoodPreference	Bill	\
0	1	20-25	4	Ibis	Veg	1300
1	2	30-35	5	LemonTree	Non-Veg	2000
2	3	25-30	6	RedFox	Veg	1322
3	4	20-25	-1	LemonTree	Veg	1234
4	5	35+	3	Ibis	Veg	989
5	6	35+	3	Ibys	Non-Veg	1909
6	7	35+	4	RedFox	Veg	1000
7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	non-Veg	-6755

NoOfPax	EstimatedSalary	Age_Group.1
0	40000	20-25
1	59000	30-35
2	30000	25-30
3	120000	20-25
4	45000	35+
5	122220	35+
6	21122	35+
7	345673	20-25
8	-99999	25-30
10	87777	30-35

```
[227]: df1['Hotel']=df1['Hotel'].str.replace('Ibys', 'Ibis')
```

```
[228]: df1
```

```
[228]:   CustomerID Age_Group Rating(1-5) Hotel FoodPreference Bill \
0           1    20-25          4     Ibis      Veg  1300
1           2    30-35          5 LemonTree Non-Veg 2000
2           3    25-30          6 RedFox      Veg 1322
3           4    20-25         -1 LemonTree      Veg 1234
4           5     35+           3     Ibis      Veg  989
5           6     35+           3     Ibis Non-Veg 1909
6           7     35+           4 RedFox      Veg 1000
7           8    20-25          7 LemonTree      Veg 2999
8           9    25-30          2     Ibis Non-Veg 3456
10          10   30-35          5 RedFox non-Veg -6755
```

```
NoOfPax EstimatedSalary Age_Group.1
0           2        40000    20-25
1           3        59000    30-35
2           2        30000    25-30
3           2       120000    20-25
4           2        45000    35+
5           2      122220    35+
6          -1        21122    35+
7          -10      345673    20-25
8           3       -99999    25-30
10          4        87777    30-35
```

```
[229]: condition = df1['Bill'] < 0
print(condition)
```

```
0    False
1    False
2    False
3    False
4    False
5    False
6    False
7    False
8    False
10   True
Name: Bill, dtype: bool
```

```
[230]: df1.loc[condition, 'Bill'] = 0
```

```
[231]: df1
```

```
[231]:   CustomerID Age_Group Rating(1-5) Hotel FoodPreference Bill \
0           1    20-25          4     Ibis      Veg  1300
1           2    30-35          5 LemonTree Non-Veg 2000
2           3    25-30          6 RedFox      Veg 1322
3           4    20-25         -1 LemonTree      Veg 1234
```

4	5	35+	3	Ibis	Veg	989
5	6	35+	3	Ibis	Non-Veg	1909
6	7	35+	4	RedFox	Veg	1000
7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	non-Veg	0

NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000
1	3	59000
2	2	30000
3	2	120000
4	2	45000
5	2	122220
6	-1	21122
7	-10	345673
8	3	-999999
10	4	87777

```
[232]: condition = df1['EstimatedSalary'] < 0
```

```
[235]: df1.loc[condition, 'EstimatedSalary'] = 0
```

```
[236]: df1
```

CustomerID	Age_Group	Rating(1-5)	Hotel	FoodPreference	Bill	\
0	1	20-25	4	Ibis	Veg	1300
1	2	30-35	5	LemonTree	Non-Veg	2000
2	3	25-30	6	RedFox	Veg	1322
3	4	20-25	-1	LemonTree	Veg	1234
4	5	35+	3	Ibis	Veg	989
5	6	35+	3	Ibis	Non-Veg	1909
6	7	35+	4	RedFox	Veg	1000
7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	non-Veg	0

NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000
1	3	59000
2	2	30000
3	2	120000
4	2	45000
5	2	122220
6	-1	21122
7	-10	345673
8	3	0

```
10        4        87777      30-35
```

```
[237]: condition1=[]
condition1.append(df1['NoOfPax'] < 0)
```

```
[241]: for i in condition1:
    df1.loc[i, 'NoOfPax'] = 0
```

```
[242]: df1
```

```
[242]:   CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill \
0            1    20-25          4     Ibis       Veg    1300
1            2    30-35          5  LemonTree    Non-Veg    2000
2            3    25-30          6    RedFox       Veg    1322
3            4    20-25         -1  LemonTree       Veg    1234
4            5      35+          3     Ibis       Veg     989
5            6      35+          3     Ibis    Non-Veg    1909
6            7      35+          4    RedFox       Veg    1000
7            8    20-25          7  LemonTree       Veg    2999
8            9    25-30          2     Ibis    Non-Veg    3456
10           10    30-35          5    RedFox    non-Veg      0
```

	NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000	20-25
1	3	59000	30-35
2	2	30000	25-30
3	2	120000	20-25
4	2	45000	35+
5	2	122220	35+
6	0	21122	35+
7	0	345673	20-25
8	0	0	25-30
10	4	87777	30-35

```
[243]: condition = df1['Rating(1-5)'] < 0
```

```
[244]: df1.loc[condition, 'Rating(1-5)'] = 0
```

```
[245]: df1
```

```
[245]:   CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill \
0            1    20-25          4     Ibis       Veg    1300
1            2    30-35          5  LemonTree    Non-Veg    2000
2            3    25-30          6    RedFox       Veg    1322
3            4    20-25          0  LemonTree       Veg    1234
4            5      35+          3     Ibis       Veg     989
5            6      35+          3     Ibis    Non-Veg    1909
```

6	7	35+	4	RedFox	Veg	1000
7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	non-Veg	0

NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000 20-25
1	3	59000 30-35
2	2	30000 25-30
3	2	120000 20-25
4	2	45000 35+
5	2	122220 35+
6	0	21122 35+
7	0	345673 20-25
8	0	0 25-30
10	4	87777 30-35

```
[246]: df1['FoodPreference']=df1['FoodPreference'].str.replace('non-Veg','Non-Veg')
```

```
[247]: df1
```

CustomerID	Age_Group	Rating(1-5)	Hotel	FoodPreference	Bill	\
0	1	20-25	4	Ibis	Veg	1300
1	2	30-35	5	LemonTree	Non-Veg	2000
2	3	25-30	6	RedFox	Veg	1322
3	4	20-25	0	LemonTree	Veg	1234
4	5	35+	3	Ibis	Veg	989
5	6	35+	3	Ibis	Non-Veg	1909
6	7	35+	4	RedFox	Veg	1000
7	8	20-25	7	LemonTree	Veg	2999
8	9	25-30	2	Ibis	Non-Veg	3456
10	10	30-35	5	RedFox	Non-Veg	0

NoOfPax	EstimatedSalary	Age_Group.1
0	2	40000 20-25
1	3	59000 30-35
2	2	30000 25-30
3	2	120000 20-25
4	2	45000 35+
5	2	122220 35+
6	0	21122 35+
7	0	345673 20-25
8	0	0 25-30
10	4	87777 30-35

```
[248]: condition = df1['Rating(1-5)'] == 0
```

```
[261]: df1.loc[condition, 'Rating(1-5)'] = round(df1['Rating(1-5)'].mean(), 2)
```

```
[262]: df1
```

```
[262]:   CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill  \
0           1    20-25       4.00     Ibis        Veg  1300
1           2    30-35       5.00  LemonTree      Non-Veg  2000
2           3    25-30       6.00     RedFox        Veg  1322
3           4    20-25      4.33  LemonTree        Veg  1234
4           5     35+       3.00     Ibis        Veg  989
5           6     35+       3.00     Ibis      Non-Veg  1909
6           7     35+       4.00     RedFox        Veg  1000
7           8    20-25       7.00  LemonTree        Veg  2999
8           9    25-30       2.00     Ibis      Non-Veg  3456
10          10    30-35       5.00     RedFox      Non-Veg    0

   NoOfPax  EstimatedSalary  Age_Group.1
0         2            40000    20-25
1         3            59000    30-35
2         2            30000    25-30
3         2            120000   20-25
4         2            45000     35+
5         2           122220    35+
6         0            21122     35+
7         0            345673   20-25
8         0              0    25-30
10        4            87777    30-35
```

```
[265]: condition = df1['Bill'] == 0
```

```
[266]: df1.loc[condition, 'Bill'] = round(df1['Bill'].mean(), 2)
```

```
[267]: df1
```

```
[267]:   CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill  \
0           1    20-25       4.00     Ibis        Veg  1300.0
1           2    30-35       5.00  LemonTree      Non-Veg  2000.0
2           3    25-30       6.00     RedFox        Veg  1322.0
3           4    20-25      4.33  LemonTree        Veg  1234.0
4           5     35+       3.00     Ibis        Veg  989.0
5           6     35+       3.00     Ibis      Non-Veg  1909.0
6           7     35+       4.00     RedFox        Veg  1000.0
7           8    20-25       7.00  LemonTree        Veg  2999.0
8           9    25-30       2.00     Ibis      Non-Veg  3456.0
10          10    30-35       5.00     RedFox      Non-Veg  1620.9
```

```
NoOfPax  EstimatedSalary  Age_Group.1
```

```

0      2      40000    20-25
1      3      59000    30-35
2      2      30000    25-30
3      2     120000   20-25
4      2      45000    35+
5      2    122220    35+
6      0     21122    35+
7      0    345673   20-25
8      0        0    25-30
10     4     87777    30-35

```

```
[270]: condition1=[]
condition1.append(df1['NoOfPax'] == 0)
```

```
[276]: for i in condition1:
    df1.loc[i, 'NoOfPax'] = df1['NoOfPax'].mean()
```

```
[277]: df1
```

```
[277]: CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill \
0          1    20-25       4.00    Ibis      Veg  1300.0
1          2    30-35       5.00  LemonTree  Non-Veg  2000.0
2          3    25-30       6.00   RedFox      Veg  1322.0
3          4    20-25       4.33  LemonTree      Veg  1234.0
4          5     35+       3.00    Ibis      Veg   989.0
5          6     35+       3.00    Ibis  Non-Veg  1909.0
6          7     35+       4.00   RedFox      Veg  1000.0
7          8    20-25       7.00  LemonTree      Veg  2999.0
8          9    25-30       2.00    Ibis  Non-Veg  3456.0
10         10   30-35       5.00   RedFox  Non-Veg  1620.9
```

	NoOfPax	EstimatedSalary	Age_Group.1
0	2.0	40000	20-25
1	3.0	59000	30-35
2	2.0	30000	25-30
3	2.0	120000	20-25
4	2.0	45000	35+
5	2.0	122220	35+
6	2.3	21122	35+
7	2.3	345673	20-25
8	2.3	0	25-30
10	4.0	87777	30-35

```
[280]: df1=df1.drop('Age_Group.1',axis=1)
```

```
[282]: condition = df1['EstimatedSalary'] == 0
```

```
[288]: df1.loc[condition, 'EstimatedSalary'] = int(round(df1['EstimatedSalary'].  
       ↪median(),0))
```

```
[289]: df1
```

```
[289]:   CustomerID  Age_Group  Rating(1-5)  Hotel  FoodPreference  Bill  \\\n      0           1    20-25        4.00    Ibis        Veg  1300.0\n      1           2    30-35        5.00  LemonTree    Non-Veg  2000.0\n      2           3    25-30        6.00  RedFox        Veg  1322.0\n      3           4    20-25        4.33  LemonTree        Veg  1234.0\n      4           5     35+        3.00    Ibis        Veg   989.0\n      5           6     35+        3.00    Ibis    Non-Veg  1909.0\n      6           7     35+        4.00  RedFox        Veg  1000.0\n      7           8    20-25        7.00  LemonTree        Veg  2999.0\n      8           9    25-30        2.00    Ibis    Non-Veg  3456.0\n      10          10   30-35        5.00  RedFox    Non-Veg  1620.9\n\n      NoOfPax  EstimatedSalary\n      0         2.0        40000.0\n      1         3.0        59000.0\n      2         2.0        30000.0\n      3         2.0       120000.0\n      4         2.0        45000.0\n      5         2.0       122220.0\n      6         2.3        21122.0\n      7         2.3       345673.0\n      8         2.3        62510.0\n      10        4.0        87777.0
```

```
[293]: df1.to_csv('Hotel_Dataset.csv',index=False)
```

```
[ ]: #create an own csv file in pandas
```

```
[314]: import pandas as pd  
import numpy as np
```

```
[315]: df2=pd.read_csv('bookstore_inventory.csv')
```

```
[316]: df2
```

```
[316]:   Book_ID          Title          Author  \\\n      0      101  The Great Gatsby  F. Scott Fitzgerald\n      1      102  To Kill a Mockingbird  Harper Lee\n      2      103            1984  George Orwell\n      3      104  Pride and Prejudice  Jane Austen\n      4      105  The Catcher in the Rye  J. D. Salinger\n      5      106        The Hobbit  J. R. R. Tolkien
```

6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
10	111	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher \
0	Fiction	10.99	25.0	Scribner
1	Fiction	8.5	NaN	J. B. Lippincott & Co
2	Dystopian	9.99	40.0	Secker & Warburg
3	Romance	abc	30.0	T. Egerton
4	Fiction	7.5	20.0	Little, Brown and Company
5	Fantasy	12.75	50.0	George Allen & Unwin
6	Fantasy	11.2	100.0	Bloomsbury
7	Fantasy	15	NaN	George Allen & Unwin
8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
10	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriller	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	20.0	The Russian Messenger
19	Epic Poetry	10	NaN	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH

```

9          1988    English
10         1988    English
11         2003    english
12         2013    English
13         2000    ENglish
14         2003    English
15         1970    Spansih
16         1866    Russian
17         1851    English
18         1869    English
19        -800     Greek

```

[317]: df2.drop\_duplicates(inplace=True)

[318]: df2

	Book_ID	Title	Author	\
0	101	The Great Gatsby	F. Scott Fitzgerald	
1	102	To Kill a Mockingbird	Harper Lee	
2	103	1984	George Orwell	
3	104	Pride and Prejudice	Jane Austen	
4	105	The Catcher in the Rye	J. D. Salinger	
5	106	The Hobbit	J. R. R. Tolkien	
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling	
7	108	The Lord of the Rings	J. R. R. Tolkien	
8	109	Animal Farm	George Orwell	
9	110	The Alchemist	Paulo Coelho	
10	111	The Alchemist	Paulo Coelho	
11	112	The Davinci Code	Dan Brown	
12	113	Inferno	Dan Brown	
13	114	Angels & Demons	Dan Brown	
14	115	The Kite Runner	Khaled Hosseini	
15	116	One Hundred Years of Solitude	Gabriel García Márquez	
16	117	Crime and Punishment	Fyodor Dostoevsky	
17	118	Moby Dick	Herman Melville	
18	119	War and Peace	Leo Tolstoy	
19	120	The Odyssey	Homer	

	Genre	Price	Stock	Publisher	\
0	Fiction	10.99	25.0	Scribner	
1	Fiction	8.5	NaN	J. B. Lippincott & Co	
2	Dystopian	9.99	40.0	Secker & Warburg	
3	Romance	abc	30.0	T. Egerton	
4	Fiction	7.5	20.0	Little, Brown and Company	
5	Fantasy	12.75	50.0	George Allen & Unwin	
6	Fantasy	11.2	100.0	Bloomsburry	
7	Fantasy	15	NaN	George Allen & Unwin	

8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
10	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriler	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	20.0	The Russian Messenger
19	Epic Poetry	10	NaN	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
10	1988	English
11	2003	english
12	2013	English
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	Englsh
19	-800	Greek

```
[319]: df2.drop_duplicates(subset=[col for col in df.columns if col != 'Book_ID'],  
↪inplace=True)
```

```
[320]: df2
```

	Book_ID	Title	Author	\
0	101	The Great Gatsby	F. Scott Fitzgerald	
1	102	To Kill a Mockingbird	Harper Lee	
2	103	1984	George Orwell	
3	104	Pride and Prejudice	Jane Austen	
4	105	The Catcher in the Rye	J. D. Salinger	
5	106	The Hobbit	J. R. R. Tolkien	

6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
10	111	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher \
0	Fiction	10.99	25.0	Scribner
1	Fiction	8.5	NaN	J. B. Lippincott & Co
2	Dystopian	9.99	40.0	Secker & Warburg
3	Romance	abc	30.0	T. Egerton
4	Fiction	7.5	20.0	Little, Brown and Company
5	Fantasy	12.75	50.0	George Allen & Unwin
6	Fantasy	11.2	100.0	Bloomsbury
7	Fantasy	15	NaN	George Allen & Unwin
8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
10	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriller	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	20.0	The Russian Messenger
19	Epic Poetry	10	NaN	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH

```

9          1988    English
10         1988  English
11          2003   english
12          2013    English
13          2000   ENglish
14          2003    English
15          1970   Spansih
16          1866    Russian
17          1851    English
18          1869    English
19          -800     Greek

```

```
[321]: for col in df2.columns:
    if df2[col].dtype == 'object':
        df2[col] = df2[col].str.strip()
    elif col != 'Book_ID':
        df2[col] = df2[col].astype(str)
```

```
[322]: df2 = df2.drop_duplicates(subset=[i for i in df2.columns if i != 'Book_ID'])
```

```
[323]: df2
```

	Book_ID	Title	Author	\	
0	101	The Great Gatsby	F. Scott Fitzgerald		
1	102	To Kill a Mockingbird	Harper Lee		
2	103	1984	George Orwell		
3	104	Pride and Prejudice	Jane Austen		
4	105	The Catcher in the Rye	J. D. Salinger		
5	106	The Hobbit	J. R. R. Tolkien		
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling		
7	108	The Lord of the Rings	J. R. R. Tolkien		
8	109	Animal Farm	George Orwell		
9	110	The Alchemist	Paulo Coelho		
11	112	The Davinci Code	Dan Brown		
12	113	Inferno	Dan Brown		
13	114	Angels & Demons	Dan Brown		
14	115	The Kite Runner	Khaled Hosseini		
15	116	One Hundred Years of Solitude	Gabriel García Márquez		
16	117	Crime and Punishment	Fyodor Dostoevsky		
17	118	Moby Dick	Herman Melville		
18	119	War and Peace	Leo Tolstoy		
19	120	The Odyssey	Homer		
	Genre	Price	Stock	Publisher	\
0	Fiction	10.99	25.0	Scribner	
1	Fiction	8.5	nan	J. B. Lippincott & Co	
2	Dystopian	9.99	40.0	Secker & Warburg	

3	Romance	abc	30.0	T. Egerton
4	Fiction	7.5	20.0	Little, Brown and Company
5	Fantasy	12.75	50.0	George Allen & Unwin
6	Fantasy	11.2	100.0	Bloomsburry
7	Fantasy	15	nan	George Allen & Unwin
8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriler	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	20.0	The Russian Messenger
19	Epic Poetry	10	nan	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english
12	2013	English
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	English
19	-800	Greek

```
[324]: column=[]
```

```
[325]: for i in range(len(df2['Stock'])):
    if(df2['Stock'].iloc[i]=='nan'):
        column.append(i)
```

```
[326]: column
```

```
[326]: [1, 7, 18]
```

```
[327]: df2['Stock'].info()

<class 'pandas.core.series.Series'>
Index: 19 entries, 0 to 19
Series name: Stock
Non-Null Count Dtype
-----
19 non-null    object
dtypes: object(1)
memory usage: 304.0+ bytes
```

```
[328]: df2['Stock'].info()

<class 'pandas.core.series.Series'>
Index: 19 entries, 0 to 19
Series name: Stock
Non-Null Count Dtype
-----
19 non-null    object
dtypes: object(1)
memory usage: 304.0+ bytes
```

```
[329]: for i in column:
         df2.loc[i, 'Stock'] = 0
```

```
[330]: df2
```

	Book_ID	Title	Author
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher	\
0	Fiction	10.99	25.0	Scribner	
1	Fiction	8.5	0	J. B. Lippincott & Co	
2	Dystopian	9.99	40.0	Secker & Warburg	
3	Romance	abc	30.0	T. Egerton	
4	Fiction	7.5	20.0	Little, Brown and Company	
5	Fantasy	12.75	50.0	George Allen & Unwin	
6	Fantasy	11.2	100.0	Bloomsburry	
7	Fantasy	15	0	George Allen & Unwin	
8	Satire	5.99	70.0	Secker & Warburg	
9	Adventure	6.75	60.0	HarperOne	
11	Thriller	9.5	80.0	Doubleday	
12	NaN	10.25	90.0	Doubleday	
13	Thriler	8.9	85.0	Doubleday	
14	Drama	7.2	100.0	Riverhead Books	
15	Magical Realism	13.3	45.0	Harper & Row	
16	Philosophical	11.5	55.0	The Russian Messenger	
17	Adventure	not available	30.0	Harper & Brothers	
18	Historical	14	0	The Russian Messenger	
19	Epic Poetry	10	nan	Ancient Greece Publishing	

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english
12	2013	English
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	Englsh
19	-800	Greek

```
[331]: df2.loc[:, 'Stock'] = df2['Stock'].fillna(0)
```

```
[332]: df2
```

[332] :	Book_ID	Title	Author \
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher \
0	Fiction	10.99	25.0	Scribner
1	Fiction	8.5	0	J. B. Lippincott & Co
2	Dystopian	9.99	40.0	Secker & Warburg
3	Romance	abc	30.0	T. Egerton
4	Fiction	7.5	20.0	Little, Brown and Company
5	Fantasy	12.75	50.0	George Allen & Unwin
6	Fantasy	11.2	100.0	Bloomsburry
7	Fantasy	15	0	George Allen & Unwin
8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriler	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	0	The Russian Messenger
19	Epic Poetry	10	nan	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english

```

4          1951      EN
5          1937      Eng
6          1997  English
7          1954  English
8          1945 ENGLISH
9          1988  English
11         2003  english
12         2013  English
13         2000 ENglish
14         2003  English
15         1970  Spansih
16         1866  Russian
17         1851  English
18         1869  English
19        -800   Greek

```

```
[333]: value = df2['Stock'].median()
```

```
[334]: df2.loc[df2['Stock'] == 0.0, 'Stock'] = value
```

```
[335]: df2
```

	Book_ID	Title	Author	\	
0	101	The Great Gatsby	F. Scott Fitzgerald		
1	102	To Kill a Mockingbird	Harper Lee		
2	103	1984	George Orwell		
3	104	Pride and Prejudice	Jane Austen		
4	105	The Catcher in the Rye	J. D. Salinger		
5	106	The Hobbit	J. R. R. Tolkien		
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling		
7	108	The Lord of the Rings	J. R. R. Tolkien		
8	109	Animal Farm	George Orwell		
9	110	The Alchemist	Paulo Coelho		
11	112	The Davinci Code	Dan Brown		
12	113	Inferno	Dan Brown		
13	114	Angels & Demons	Dan Brown		
14	115	The Kite Runner	Khaled Hosseini		
15	116	One Hundred Years of Solitude	Gabriel García Márquez		
16	117	Crime and Punishment	Fyodor Dostoevsky		
17	118	Moby Dick	Herman Melville		
18	119	War and Peace	Leo Tolstoy		
19	120	The Odyssey	Homer		
	Genre	Price	Stock	Publisher	\
0	Fiction	10.99	25.0	Scribner	
1	Fiction	8.5	47.5	J. B. Lippincott & Co	
2	Dystopian	9.99	40.0	Secker & Warburg	

3	Romance	abc	30.0	T. Egerton
4	Fiction	7.5	20.0	Little, Brown and Company
5	Fantasy	12.75	50.0	George Allen & Unwin
6	Fantasy	11.2	100.0	Bloomsburry
7	Fantasy	15	47.5	George Allen & Unwin
8	Satire	5.99	70.0	Secker & Warburg
9	Adventure	6.75	60.0	HarperOne
11	Thriller	9.5	80.0	Doubleday
12	NaN	10.25	90.0	Doubleday
13	Thriler	8.9	85.0	Doubleday
14	Drama	7.2	100.0	Riverhead Books
15	Magical Realism	13.3	45.0	Harper & Row
16	Philosophical	11.5	55.0	The Russian Messenger
17	Adventure	not available	30.0	Harper & Brothers
18	Historical	14	47.5	The Russian Messenger
19	Epic Poetry	10	nan	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english
12	2013	English
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	Englsh
19	-800	Greek

```
[336]: df2['Stock'].astype(float)
```

```
[336]: 0    25.0
1    47.5
2    40.0
3    30.0
4    20.0
5    50.0
6   100.0
```

```
7      47.5
8      70.0
9      60.0
11     80.0
12     90.0
13     85.0
14    100.0
15     45.0
16     55.0
17     30.0
18     47.5
19      NaN
Name: Stock, dtype: float64
```

```
[337]: df2 = df.copy()
df2['Stock'].fillna(value)
```

```
[337]: 0      46.315789
1      46.315789
2      40.000000
3      30.000000
4      20.000000
5      50.000000
6      100.000000
7      46.315789
8      70.000000
9      60.000000
11     80.000000
12     90.000000
13     85.000000
14    100.000000
15     45.000000
16     55.000000
17     30.000000
18      0.000000
19      0.000000
Name: Stock, dtype: float64
```

```
[338]: df2
```

```
[338]:   Book_ID          Title           Author \
0      101      The Great Gatsby  F. Scott Fitzgerald
1      102  To Kill a Mockingbird  Harper Lee
2      103            1984        George Orwell
3      104  Pride and Prejudice  Jane Austen
4      105  The Catcher in the Rye  J. D. Salinger
5      106        The Hobbit  J. R. R. Tolkien
```

6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher \
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday
12	NaN	10	90.000000	Doubleday
13	Thriller	8	85.000000	Doubleday
14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger
19	Epic Poetry	10	0.000000	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english

```
12          2013 English
13          2000 ENglish
14          2003 English
15          1970 Spansih
16          1866 Russian
17          1851 English
18          1869 English
19          -800 Greek
```

```
[339]: df2['Price'].info()
```

```
<class 'pandas.core.series.Series'>
Index: 19 entries, 0 to 19
Series name: Price
Non-Null Count Dtype
-----
19 non-null    int64
dtypes: int64(1)
memory usage: 860.0+ bytes
```

```
[340]: df2['Price'].info()
```

```
<class 'pandas.core.series.Series'>
Index: 19 entries, 0 to 19
Series name: Price
Non-Null Count Dtype
-----
19 non-null    int64
dtypes: int64(1)
memory usage: 860.0+ bytes
```

```
[341]: df2['Price'].fillna(0)
```

```
[341]: 0      10
1      8
2      9
3      8
4      7
5      12
6      11
7      15
8      5
9      6
11     9
12     10
13     8
14     7
15     13
```

```

16    11
17     8
18    14
19    10
Name: Price, dtype: int64

```

```
[342]: val=df2['Price'].mean()
```

```
[343]: for i in range(len(df2['Price'])):
    if df2.iloc[i]['Price'] == 0:
        df2.iloc[i]['Price'] =val
```

```
[344]: df2
```

	Book_ID	Title	Author
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy
19	120	The Odyssey	Homer

	Genre	Price	Stock	Publisher
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday

12	NaN	10	90.000000	Doubleday
13	Thriler	8	85.000000	Doubleday
14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger
19	Epic Poetry	10	0.000000	Ancient Greece Publishing

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english
12	2013	English
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	English
19	-800	Greek

```
[345]: df2['Year_Published']=df2['Year_Published'].astype(int)
```

```
[346]: df2['Year_Published'].info()
```

```
<class 'pandas.core.series.Series'>
Index: 19 entries, 0 to 19
Series name: Year_Published
Non-Null Count Dtype
-----
19 non-null    int64
dtypes: int64(1)
memory usage: 860.0+ bytes
```

```
[347]: df2 = df2[df2['Year_Published'] >= 0]
```

```
[348]: df2
```

[348] :	Book_ID	Title	Author \
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy

	Genre	Price	Stock	Publisher \
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday
12	NaN	10	90.000000	Doubleday
13	Thriller	8	85.000000	Doubleday
14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng

```

6          1997 English
7          1954 English
8          1945 ENGLISH
9          1988 English
11         2003 english
12         2013 English
13         2000 ENglish
14         2003 English
15         1970 Spansih
16         1866 Russian
17         1851 English
18         1869 Englsh

```

[349]: df2.dropna()

	Book_ID	Title	Author
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy

	Genre	Price	Stock	Publisher
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday
13	Thriller	8	85.000000	Doubleday

14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger

	Year_Published	Language
0	1925	English
1	1960	ENGLISH
2	1949	English
3	1813	english
4	1951	EN
5	1937	Eng
6	1997	English
7	1954	English
8	1945	ENGLISH
9	1988	English
11	2003	english
13	2000	ENglish
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	Englsh

```
[353]: df2 = df2.copy()
df2['Language']=df2['Language'].str.
    .replace(r'\b(ENGLISH|EN|Eng|english|ENglish|Englsh)\b', 'English', regex=True)
```

```
[354]: df2
```

	Book_ID	Title	Author
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez

16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy

	Genre	Price	Stock	Publisher \
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday
12	NaN	10	90.000000	Doubleday
13	Thriller	8	85.000000	Doubleday
14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger

	Year_Published	Language
0	1925	English
1	1960	English
2	1949	English
3	1813	English
4	1951	English
5	1937	English
6	1997	English
7	1954	English
8	1945	English
9	1988	English
11	2003	English
12	2013	English
13	2000	English
14	2003	English
15	1970	Spansih
16	1866	Russian
17	1851	English
18	1869	English

[352] : df2

[352] :	Book_ID	Title	Author \
0	101	The Great Gatsby	F. Scott Fitzgerald
1	102	To Kill a Mockingbird	Harper Lee
2	103	1984	George Orwell
3	104	Pride and Prejudice	Jane Austen
4	105	The Catcher in the Rye	J. D. Salinger
5	106	The Hobbit	J. R. R. Tolkien
6	107	Harry Potter and the Philosopher's Stone	J.K. Rowling
7	108	The Lord of the Rings	J. R. R. Tolkien
8	109	Animal Farm	George Orwell
9	110	The Alchemist	Paulo Coelho
11	112	The Davinci Code	Dan Brown
12	113	Inferno	Dan Brown
13	114	Angels & Demons	Dan Brown
14	115	The Kite Runner	Khaled Hosseini
15	116	One Hundred Years of Solitude	Gabriel García Márquez
16	117	Crime and Punishment	Fyodor Dostoevsky
17	118	Moby Dick	Herman Melville
18	119	War and Peace	Leo Tolstoy

	Genre	Price	Stock	Publisher \
0	Fiction	10	46.315789	Scribner
1	Fiction	8	46.315789	J. B. Lippincott & Co
2	Dystopian	9	40.000000	Secker & Warburg
3	Romance	8	30.000000	T. Egerton
4	Fiction	7	20.000000	Little, Brown and Company
5	Fantasy	12	50.000000	George Allen & Unwin
6	Fantasy	11	100.000000	Bloomsbury
7	Fantasy	15	46.315789	George Allen & Unwin
8	Satire	5	70.000000	Secker & Warburg
9	Adventure	6	60.000000	HarperOne
11	Thriller	9	80.000000	Doubleday
12	NaN	10	90.000000	Doubleday
13	Thriller	8	85.000000	Doubleday
14	Drama	7	100.000000	Riverhead Books
15	Magical Realism	13	45.000000	Harper & Row
16	Philosophical	11	55.000000	The Russian Messenger
17	Adventure	8	30.000000	Harper & Brothers
18	Historical	14	0.000000	The Russian Messenger

	Year_Published	Language
0	1925	English
1	1960	English
2	1949	English
3	1813	English
4	1951	English
5	1937	English

```
6      1997 English
7      1954 English
8      1945 English
9      1988 English
11     2003 English
12     2013 English
13     2000 English
14     2003 English
15     1970 Spansih
16     1866 Russian
17     1851 English
18     1869 English
```

```
[355]: df2.to_csv('bookstore_inventory.csv')
```