
```
import pandas as pd
import numpy as np

df = pd.read_csv(r'C:\Users\hp\Desktop\100DaysOfDataScience\Day 22\
pandasrevision.csv')
df.head()
```

```
   Emp_No             Emp_Name  Emp_Age  Working Experience In This
Firm  \
0    1001            Amy White     70.0
5.0
1    1002         Jack Stewart     53.0
8.0
2    1003  Richard Lauderdale     35.0
3.0
3    1004         Sara Johnson     43.0
10.0
4    1005       Shayla Shields     24.0
NaN

  Department         Country
0         IT         Seattle
1      Admin        New york
2      Admin  San Francisco
3         HR         Chicago
4    Finance          Dallas
```

```
df.shape
```

```
(200, 6)
```

```
df.columns
```

```
Index(['Emp_No', 'Emp_Name', 'Emp_Age', 'Working Experience In This
Firm',
       'Department', 'Country'],
      dtype='object')
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 6 columns):
 #   Column                           Non-Null Count  Dtype
---  ------                           --------------  -----
 0   Emp_No                           200 non-null    int64
 1   Emp_Name                         200 non-null    object
 2   Emp_Age                          159 non-null    float64
 3   Working Experience In This Firm  171 non-null    float64
 4   Department                       198 non-null    object
 5   Country                          200 non-null    object
dtypes: float64(2), int64(1), object(3)
memory usage: 9.5+ KB
```

```
df.describe()
```

```
            Emp_No     Emp_Age  Working Experience In This Firm
count   200.000000  159.000000                       171.000000
mean   1100.500000   39.408805                         8.298246
std      57.879185   13.636425                         6.045800
min    1001.000000   18.000000                         1.000000
25%    1050.750000   29.000000                         3.000000
50%    1100.500000   37.000000                         8.000000
75%    1150.250000   49.000000                        11.000000
max    1200.000000   70.000000                        29.000000
```

```
df.nunique()
```

```
Emp_No                             200
Emp_Name                           200
Emp_Age                             46
Working Experience In This Firm     24
Department                           7
Country                              5
dtype: int64
```

```
df1 = df.copy()
df1 = df1.rename(columns={"Emp_No":"Employee
Number","Emp_Name":"Name","Emp_Age":"Age",
                          "Country":"City","Working Experience In This
Firm":"Experience"})
df1.columns
```

```
Index(['Employee Number', 'Name', 'Age', 'Experience', 'Department',
'City'], dtype='object')
```

```
df.tail()
```

```
     Emp_No      Emp_Name  Emp_Age  Working Experience In This Firm  \
195    1196  Adriana Rios      NaN                               1.0
196    1197     Emma Mann     21.0                               5.0
197    1198  Lacie Gordon     29.0                               NaN
198    1199   Cara Hilton     65.0                              29.0
199    1200  Francis Horn     57.0                              14.0

    Department        Country
195      Sales         Dallas
196      Sales  San Francisco
197      Admin       New york
198         IT        Seattle
199    Finance  San Francisco
```

```
df1.columns = df1.columns.str.replace(" ","")
df1.head(10)
```

```
   EmployeeNumber                Name   Age  Experience Department  \
0            1001           Amy White  70.0         5.0         IT
1            1002        Jack Stewart  53.0         8.0      Admin
2            1003  Richard Lauderdale  35.0         3.0      Admin
3            1004        Sara Johnson  43.0        10.0         HR
4            1005      Shayla Shields  24.0         NaN    Finance
5            1006       Milly Cordova  29.0         2.0  Marketing
6            1007         Greta Poole   NaN        15.0         HR
7            1008       Mustafa Welsh   NaN        10.0  Marketing
8            1009       Richard Gibbs  47.0         3.0         IT
9            1010          Neo Mcneil  43.0         7.0  Marketing

            City
0        Seattle
1       New york
2  San Francisco
3        Chicago
4         Dallas
5        Seattle
6        Seattle
7        Seattle
8        Seattle
9        Seattle
```

```
print(df1.duplicated().sum())
print(df1.shape)
```

```
0
(200, 6)
```

```
df1.isna().sum()
```

```
EmployeeNumber     0
Name               0
Age               41
Experience        29
Department         2
City               0
dtype: int64
```

```
df2 = df1.copy()
```

```
df2 = df2.dropna(subset=['Department'])
print(df2.isna().sum())
print('*' * 100)
print("Shape: ",df2.shape)
```

```
EmployeeNumber     0
Name               0
Age               41
Experience        29
Department         0
City               0
dtype: int64
****************************************************************************************************
Shape:  (198, 6)
```

```
print('*' * 100)
print("Unique: ",df2['Age'].unique())
print('*' * 100)
print("Mean: ",df2['Age'].mean())
print('*' * 100)
print("Median: ",df2['Age'].median())
print('*' * 100)
print("Mode: ",df2['Age'].mode())
print('*' * 100)
print("Description: \n",df2['Age'].describe())
print('*' * 100)
```

```
****************************************************************************************************
Unique:  [70. 53. 35. 43. 24. 29. nan 47. 19. 28. 37. 59. 54. 32. 38.
49. 51. 21.
 67. 30. 34. 27. 39. 40. 58. 36. 63. 18. 57. 44. 61. 50. 25. 46. 48.
41.
 22. 20. 60. 42. 26. 45. 31. 55. 56. 33. 65.]
****************************************************************************************************
Mean:  39.445859872611464
****************************************************************************************************
```

```
******************************
Median:  37.0
********************************************************************************
******************************
Mode:  0    37.0
Name: Age, dtype: float64
********************************************************************************
******************************
Description:
 count    157.000000
mean      39.445860
std       13.677323
min       18.000000
25%       29.000000
50%       37.000000
75%       49.000000
max       70.000000
Name: Age, dtype: float64
********************************************************************************
******************************
```

```
df2['Age'].fillna(df2['Age'].median(), inplace = True)
print(df2.isnull().sum())
print('*' * 100)
print("Shape: ",df2.shape)
```

```
EmployeeNumber     0
Name               0
Age                0
Experience        29
Department         0
City               0
dtype: int64
********************************************************************************
******************************
Shape:  (198, 6)
```

```
print('*' * 100)
print("Unique: ",df2['Experience'].unique())
print('*' * 100)
print("Mean: ",df2['Experience'].mean())
print('*' * 100)
print("Median: ",df2['Experience'].median())
print('*' * 100)
print("Mode: ",df2['Experience'].mode())
print('*' * 100)
print("Description: \n",df2['Experience'].describe())
print('*' * 100)
```

```
**********************************************************************
******************************
Unique:  [ 5.  8.  3. 10. nan  2. 15.  7.  9. 17. 13.  4.  1. 25. 24.
29.  6. 11.
 21. 23. 16. 12. 19. 14. 27.]
**********************************************************************
******************************
Mean:  8.266272189349113
**********************************************************************
******************************
Median:  8.0
**********************************************************************
******************************
Mode:  0    7.0
Name: Experience, dtype: float64
**********************************************************************
******************************
Description:
 count    169.000000
mean       8.266272
std        6.065622
min        1.000000
25%        3.000000
50%        8.000000
75%       11.000000
max       29.000000
Name: Experience, dtype: float64
**********************************************************************
******************************
```

```
df2['Experience'].fillna(df2['Experience'].median(), inplace = True)
print(df2.isnull().sum())
print('*' * 100)
print("Shape: ",df2.shape)
```

```
EmployeeNumber    0
Name              0
Age               0
Experience        0
Department        0
City              0
dtype: int64
**********************************************************************
******************************
Shape:  (198, 6)
```

```
Experience_categories = df2['Experience']
Experience_categories = Experience_categories.dropna()

bins = [0,4,13,18,25,Experience_categories.max()]
```

```
binlabels = ['Entry Level', 'Associate', 'Manager', 'Director',
'ShareHolder']

categories = pd.cut(Experience_categories,bins,labels = binlabels)
categories.tail(20)
```

```
180      Associate
181      Associate
182    Entry Level
183      Associate
184      Associate
185       Director
186    Entry Level
187      Associate
188      Associate
189      Associate
190        Manager
191      Associate
192      Associate
193      Associate
194      Associate
195    Entry Level
196      Associate
197      Associate
198    ShareHolder
199        Manager
Name: Experience, dtype: category
Categories (5, object): ['Entry Level' < 'Associate' < 'Manager' <
'Director' < 'ShareHolder']
```

```
df2['ExperienceCategories'] = categories
df2.tail(15)
```

	EmployeeNumber	Name	Age	Experience	Department	\
185	1186	Joshua Booth	67.0	25.0	Finance	
186	1187	Skyla Thompson	38.0	2.0	Sales	
187	1188	Dennis Wilcox	37.0	9.0	IT	
188	1189	Keisha Singh	18.0	10.0	Admin	
189	1190	Bilal Lester	37.0	5.0	Sales	
190	1191	Jade Sherman	27.0	14.0	Finance	
191	1192	Cian Day	19.0	12.0	IT	
192	1193	Safa Sharp	37.0	7.0	HR	
193	1194	Grayson Valentine	33.0	9.0	IT	
194	1195	Maddie Williamson	24.0	10.0	Admin	
195	1196	Adriana Rios	37.0	1.0	Sales	
196	1197	Emma Mann	21.0	5.0	Sales	
197	1198	Lacie Gordon	29.0	8.0	Admin	
198	1199	Cara Hilton	65.0	29.0	IT	
199	1200	Francis Horn	57.0	14.0	Finance	

```
              City ExperienceCategories
185       New york             Director
186  San Francisco          Entry Level
187       New york            Associate
188        Seattle            Associate
189  San Francisco            Associate
190  San Francisco              Manager
191       New york            Associate
192        Seattle            Associate
193        Chicago            Associate
194       New york            Associate
195         Dallas           Entry Level
196  San Francisco            Associate
197       New york            Associate
198        Seattle           ShareHolder
199  San Francisco               Manager
```

```
df3 = df2.copy()

AgegreaterthanThan35andDepartmentasIT = df3[(df3['Age'] > 25) &
(df3['Department'] == 'IT')]
print("Shape: ",AgegreaterthanThan35andDepartmentasIT.shape)
AgegreaterthanThan35andDepartmentasIT
```

```
Shape:  (38, 7)
```

```
     EmployeeNumber               Name   Age  Experience Department  \
0              1001          Amy White  70.0         5.0         IT
8              1009      Richard Gibbs  47.0         3.0         IT
17             1018       Eugene Welch  37.0         8.0         IT
18             1019     Gabriel Patton  37.0        17.0         IT
27             1028      Safaa Hubbard  29.0         1.0         IT
28             1029       Bobby Lucero  37.0         7.0         IT
34             1035     Grace Jennings  34.0        13.0         IT
45             1046       Arron Austin  70.0        29.0         IT
49             1050     Leonardo Sharp  34.0         8.0         IT
54             1055      Terry Blevins  59.0        21.0         IT
55             1056   Jamil Valenzuela  36.0        11.0         IT
62             1063      Cormac Foster  67.0         8.0         IT
70             1071        Rhonda Pace  29.0         8.0         IT
72             1073     Xanthe Douglas  37.0         8.0         IT
73             1074        Abi Rodgers  44.0        11.0         IT
74             1075     Carlos Acevedo  61.0        10.0         IT
83             1084        Brian Curry  29.0        10.0         IT
84             1085  Bridget Nicholson  34.0         8.0         IT
101            1102          Alma John  50.0         7.0         IT
102            1103    Annie Henderson  27.0         8.0         IT
115            1116      Callum Duncan  37.0        11.0         IT
124            1125       Elsie Parker  47.0         7.0         IT
132            1133    Larissa Acevedo  37.0         7.0         IT
```

140	1141	Elinor Stevens	37.0	17.0	IT
141	1142	Jean Bradley	54.0	8.0	IT
144	1145	Samantha Zuniga	26.0	7.0	IT
145	1146	Subhan Sims	37.0	7.0	IT
151	1152	Edwin Rodriguez	37.0	14.0	IT
159	1160	Haleema Powers	37.0	2.0	IT
160	1161	Oskar Aguirre	41.0	7.0	IT
167	1168	Barnaby Cabrera	51.0	8.0	IT
169	1170	Harriett Knowles	37.0	12.0	IT
172	1173	Elsa Moody	41.0	2.0	IT
179	1180	Albert Foster	57.0	21.0	IT
180	1181	Candice Curtis	59.0	6.0	IT
187	1188	Dennis Wilcox	37.0	9.0	IT
193	1194	Grayson Valentine	33.0	9.0	IT
198	1199	Cara Hilton	65.0	29.0	IT

	City	ExperienceCategories
0	Seattle	Associate
8	Seattle	Entry Level
17	San Francisco	Associate
18	San Francisco	Manager
27	Seattle	Entry Level
28	Chicago	Associate
34	Seattle	Associate
45	Dallas	ShareHolder
49	Dallas	Associate
54	New york	Director
55	Seattle	Associate
62	Chicago	Associate
70	Chicago	Associate
72	Dallas	Associate
73	Seattle	Associate
74	Seattle	Associate
83	New york	Associate
84	Seattle	Associate
101	San Francisco	Associate
102	Chicago	Associate
115	New york	Associate
124	New york	Associate
132	Chicago	Associate
140	San Francisco	Manager
141	New york	Associate
144	Seattle	Associate
145	San Francisco	Associate
151	Chicago	Manager
159	San Francisco	Entry Level
160	New york	Associate
167	Seattle	Associate
169	Chicago	Associate

```
172        Chicago          Entry Level
179  San Francisco             Director
180       New york            Associate
187       New york            Associate
193        Chicago            Associate
198        Seattle           ShareHolder
```

```
Agegreaterthan35andExpCategoryasManager = df3[(df3['Age'] > 25) &
(df3['ExperienceCategories'] == 'Manager')]
print("Shape: ",Agegreaterthan35andExpCategoryasManager.shape)
Agegreaterthan35andExpCategoryasManager
```

```
Shape:  (14, 7)
```

```
     EmployeeNumber                Name   Age  Experience   Department
\
6              1007         Greta Poole  37.0        15.0           HR

18             1019      Gabriel Patton  37.0        17.0           IT

22             1023         Kelvin Knox  38.0        15.0        Admin

76             1077      Krish Shepherd  50.0        17.0      Finance

80             1081        Louisa Lopez  61.0        16.0           HR

105            1106        Rosanna Cook  57.0        17.0    Marketing

110            1111      Jorge Mckenzie  47.0        15.0        Admin

112            1113     Emelia O'Connor  32.0        14.0  Procurement

140            1141      Elinor Stevens  37.0        17.0           IT

149            1150  SSeattlennah Ross   28.0        15.0        Admin

151            1152     Edwin Rodriguez  37.0        14.0           IT

168            1169         Frank Rivas  61.0        14.0           HR

190            1191        Jade Sherman  27.0        14.0      Finance

199            1200        Francis Horn  57.0        14.0      Finance


              City ExperienceCategories
6          Seattle              Manager
18   San Francisco              Manager
22          Dallas              Manager
76         Chicago              Manager
80        New york              Manager
```

```
105        Chicago              Manager
110       New york              Manager
112  San Francisco              Manager
140  San Francisco              Manager
149        Seattle              Manager
151        Chicago              Manager
168         Dallas              Manager
190  San Francisco              Manager
199  San Francisco              Manager
```

```
CityinandExpCategoryasDirector = df3[((df3['City'] == 'New york') |
(df3['City'] == 'San Francisco')) & (df3['ExperienceCategories'] ==
'Director')]
print("Shape: ",a.shape)
a
```

```
Shape:  (5, 7)
```

```
     EmployeeNumber            Name   Age  Experience Department  \
54             1055   Terry Blevins  59.0        21.0         IT
56             1057   Theodore Huff  67.0        23.0      Admin
111            1112  Krishan Reilly  54.0        19.0         HR
179            1180   Albert Foster  57.0        21.0         IT
185            1186    Joshua Booth  67.0        25.0    Finance

              City ExperienceCategories
54        New york             Director
56   San Francisco             Director
111       New york             Director
179  San Francisco             Director
185       New york             Director
```

```
df3.
```