

pandas

December 1, 2021

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

```
[2]: url="https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/
      ↳titanic.csv"
```

```
[3]: dataframe=pd.read_csv(url)
```

```
[4]: dataframe
```

```
[4]:
```

| | Name | PClass | Age | Sex | \ |
|------|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | |
| ... | ... | ... | ... | ... | |
| 1308 | Zakarian, Mr Artun | 3rd | 27.00 | male | |
| 1309 | Zakarian, Mr Maprieder | 3rd | 26.00 | male | |
| 1310 | Zenni, Mr Philip | 3rd | 22.00 | male | |
| 1311 | Lievens, Mr Rene | 3rd | 24.00 | male | |
| 1312 | Zimmerman, Leo | 3rd | 29.00 | male | |

| | Survived | SexCode |
|------|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |
| ... | ... | ... |
| 1308 | 0 | 0 |
| 1309 | 0 | 0 |
| 1310 | 0 | 0 |
| 1311 | 0 | 0 |
| 1312 | 0 | 0 |

[1313 rows x 6 columns]

```
[5]: dataframe.head()
```

```
[5]:
```

| | Name | PClass | Age | Sex | \ |
|---|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | |

| | Survived | SexCode |
|---|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |

```
[6]: dataframe.tail()
```

```
[6]:
```

| | Name | PClass | Age | Sex | Survived | SexCode |
|------|------------------------|--------|------|------|----------|---------|
| 1308 | Zakarian, Mr Artun | 3rd | 27.0 | male | 0 | 0 |
| 1309 | Zakarian, Mr Maprieder | 3rd | 26.0 | male | 0 | 0 |
| 1310 | Zenni, Mr Philip | 3rd | 22.0 | male | 0 | 0 |
| 1311 | Lievens, Mr Rene | 3rd | 24.0 | male | 0 | 0 |
| 1312 | Zimmerman, Leo | 3rd | 29.0 | male | 0 | 0 |

```
[7]: dataframe.head(10)
```

```
[7]:
```

| | Name | PClass | Age | Sex | \ |
|---|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | |
| 5 | Anderson, Mr Harry | 1st | 47.00 | male | |
| 6 | Andrews, Miss Kornelia Theodosia | 1st | 63.00 | female | |
| 7 | Andrews, Mr Thomas, jr | 1st | 39.00 | male | |
| 8 | Appleton, Mrs Edward Dale (Charlotte Lamson) | 1st | 58.00 | female | |
| 9 | Artagaveytia, Mr Ramon | 1st | 71.00 | male | |

| | Survived | SexCode |
|---|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |
| 5 | 1 | 0 |
| 6 | 1 | 1 |
| 7 | 0 | 0 |
| 8 | 1 | 1 |
| 9 | 0 | 0 |

```
[8]: dataframe.head(11)
```

```
[8]:
```

| | Name | PClass | Age | Sex | \ |
|----|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | |
| 5 | Anderson, Mr Harry | 1st | 47.00 | male | |
| 6 | Andrews, Miss Kornelia Theodosia | 1st | 63.00 | female | |
| 7 | Andrews, Mr Thomas, jr | 1st | 39.00 | male | |
| 8 | Appleton, Mrs Edward Dale (Charlotte Lamson) | 1st | 58.00 | female | |
| 9 | Artagaveytia, Mr Ramon | 1st | 71.00 | male | |
| 10 | Astor, Colonel John Jacob | 1st | 47.00 | male | |

| | Survived | SexCode |
|----|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |
| 5 | 1 | 0 |
| 6 | 1 | 1 |
| 7 | 0 | 0 |
| 8 | 1 | 1 |
| 9 | 0 | 0 |
| 10 | 0 | 0 |

```
[6]: datafr=pd.DataFrame()
```

```
[10]: datafr["Name"]=["harshith","varun", "vamshi", "sid", "akshay"]
```

```
[11]: datafr
```

```
[11]:
```

| | Name |
|---|----------|
| 0 | harshith |
| 1 | varun |
| 2 | vamshi |
| 3 | sid |
| 4 | akshay |

```
[9]: datafr["Age"]=[19,20,19,19,19]
datafr
```

```
[9]:
```

| | Age |
|---|-----|
| 0 | 19 |
| 1 | 20 |
| 2 | 19 |
| 3 | 19 |
| 4 | 19 |

```
[10]: new_team=pd.Series(['mukul','45'], index=['Name','Age'])
datafr.append(new_team, ignore_index=True)
```

```
[10]:   Age  Name
0   19   NaN
1   20   NaN
2   19   NaN
3   19   NaN
4   19   NaN
5   45 mukul
```

```
[8]: new_team=pd.Series(['mukul','45'], index=['Name','Age'])
datafr.append(new_team, ignore_index=True)
```

```
[8]:   Age  Name
0   19   NaN
1   20   NaN
2   19   NaN
3   19   NaN
4   19   NaN
5   45 mukul
```

```
[16]: datafr.append(new_team, ignore_index=True)
```

```
[16]:   Name Age
0 harshith 19
1    varun 20
2    vamshi 19
3     sid 19
4    akshay 19
5    mukul 45
```

```
[17]: datafr.shape
```

```
[17]: (5, 2)
```

```
[18]: datafr.describe()
```

```
[18]:           Age
count    5.000000
mean    19.200000
std      0.447214
min     19.000000
25%     19.000000
50%     19.000000
75%     19.000000
max     20.000000
```

```
[19]: datafr.iloc[0:1]
```

```
[19]:   Name Age
0 harshith 19
```

```
[20]: datafr.set_index(datafr["Age"])
```

```
[20]:      Name  Age
Age
19  harshith  19
20   varun   20
19   vamshi  19
19    sid   19
19  akshay   19
```

```
[23]: dataframe[dataframe['Sex']=='female'].head(10)
```

```
[23]:      Name PClass  Age  Sex \
0      Allen, Miss Elisabeth Walton    1st  29.0  female
1      Allison, Miss Helen Loraine    1st   2.0  female
3  Allison, Mrs Hudson JC (Bessie Waldo Daniels)    1st  25.0  female
6      Andrews, Miss Kornelia Theodosia    1st  63.0  female
8  Appleton, Mrs Edward Dale (Charlotte Lamson)    1st  58.0  female
11 Astor, Mrs John Jacob (Madeleine Talmadge Force)    1st  19.0  female
12      Aubert, Mrs Leontine Pauline    1st   NaN  female
15  Baxter, Mrs James (Helene DeLaudeniére Chaput)    1st  50.0  female
19  Beckwith, Mrs Richard Leonard (Sallie Monypeny)    1st  47.0  female
23      Bishop, Mrs Dickinson H (Helen Walton)    1st  19.0  female

      Survived  SexCode
0           1         1
1           0         1
3           0         1
6           1         1
8           1         1
11          1         1
12          1         1
15          1         1
19          1         1
23          1         1
```

```
[24]: dataframe[dataframe['Sex']=='male'].head(10)
```

```
[24]:      Name PClass  Age  Sex  Survived  SexCode
2  Allison, Mr Hudson Joshua Creighton    1st  30.00  male         0         0
4      Allison, Master Hudson Trevor    1st   0.92  male         1         0
5      Anderson, Mr Harry    1st  47.00  male         1         0
7      Andrews, Mr Thomas, jr    1st  39.00  male         0         0
9      Artagaveytia, Mr Ramon    1st  71.00  male         0         0
10     Astor, Colonel John Jacob    1st  47.00  male         0         0
13     Barkworth, Mr Algernon H    1st   NaN  male         1         0
14      Baumann, Mr John D    1st   NaN  male         0         0
16     Baxter, Mr Quigg Edmond    1st  24.00  male         0         0
17     Beattie, Mr Thomson    1st  36.00  male         0         0
```

```
[26]: dataframe[(dataframe['Sex']=='male') & (dataframe['Age']>=50)].count()
```

```
[26]: Name      48
      PClass   48
      Age      48
      Sex      48
      Survived  48
      SexCode   48
      dtype: int64
```

```
[27]: dataframe['Sex'].replace('male','man')
```

```
[27]: 0      female
      1      female
      2       man
      3      female
      4       man
      ...
     1308      man
     1309      man
     1310      man
     1311      man
     1312      man
      Name: Sex, Length: 1313, dtype: object
```

```
[28]: dataframe.rename(columns={'Survived':'ML'})
```

```
[28]:
```

| | Name | PClass | Age | Sex | ML | \ |
|------|---|--------|-------|--------|-----|----|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | 1 | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | 0 | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | 0 | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | 0 | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | 1 | |
| ... | ... | ... | ... | ... | ... | .. |
| 1308 | Zakarian, Mr Artun | 3rd | 27.00 | male | 0 | |
| 1309 | Zakarian, Mr Maprieder | 3rd | 26.00 | male | 0 | |
| 1310 | Zenni, Mr Philip | 3rd | 22.00 | male | 0 | |
| 1311 | Lievens, Mr Rene | 3rd | 24.00 | male | 0 | |
| 1312 | Zimmerman, Leo | 3rd | 29.00 | male | 0 | |

| | SexCode |
|------|---------|
| 0 | 1 |
| 1 | 1 |
| 2 | 0 |
| 3 | 1 |
| 4 | 0 |
| ... | ... |
| 1308 | 0 |
| 1309 | 0 |

```
1310      0
1311      0
1312      0
```

```
[1313 rows x 6 columns]
```

```
[1]: import pandas as pd
df=pd.DataFrame()
```

```
[2]: df["name"]=["harshith","vamshi","Varun"]
```

```
[3]: df
```

```
[3]:      name
0  harshith
1    vamshi
2    Varun
```

```
[6]: df.iloc[:,0]
```

```
[6]: 0    harshith
1      vamshi
2      Varun
Name: name, dtype: object
```

```
[7]: df["Age"]=[19,19,20]
```

```
[8]: df
```

```
[8]:      name  Age
0  harshith   19
1    vamshi   19
2    Varun   20
```

```
[11]: df["Age"].max()
```

```
[11]: 20
```

```
[12]: df["Age"].min()
```

```
[12]: 19
```

```
[15]: df["Age"].sum()
```

```
[15]: 58
```

```
[16]: df["Age"].count()
```

```
[16]: 3
```

```
[17]: df["Age"].mean()
```

```
[17]: 19.333333333333332
```

```
[13]: url="https://raw.githubusercontent.com/chrisalbon/simulated_datasets/master/
      ↪titanic.csv"
dataframe=pd.read_csv(url)
```

```
dataframe
```

```
[13]:
```

| | Name | PClass | Age | Sex | \ |
|------|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | male | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | male | |
| ... | ... | ... | ... | ... | |
| 1308 | Zakarian, Mr Artun | 3rd | 27.00 | male | |
| 1309 | Zakarian, Mr Maprieder | 3rd | 26.00 | male | |
| 1310 | Zenni, Mr Philip | 3rd | 22.00 | male | |
| 1311 | Lievens, Mr Rene | 3rd | 24.00 | male | |
| 1312 | Zimmerman, Leo | 3rd | 29.00 | male | |

| | Survived | SexCode |
|------|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |
| ... | ... | ... |
| 1308 | 0 | 0 |
| 1309 | 0 | 0 |
| 1310 | 0 | 0 |
| 1311 | 0 | 0 |
| 1312 | 0 | 0 |

```
[1313 rows x 6 columns]
```

```
[19]: dataframe["Sex"].unique()
```

```
[19]: array(['female', 'male'], dtype=object)
```

```
[20]: dataframe["Name"].unique()
```

```
[20]: array(['Allen, Miss Elisabeth Walton', 'Allison, Miss Helen Loraine',  
          'Allison, Mr Hudson Joshua Creighton', ..., 'Zenni, Mr Philip',  
          'Lievens, Mr Rene', 'Zimmerman, Leo'], dtype=object)
```

```
[21]: dataframe["Sex"].value_counts()
```

```
[21]: male      851  
female    462  
Name: Sex, dtype: int64
```

```
[22]: dataframe[dataframe["Age"].isnull()]
```

```
[22]:
```

| | Name | PClass | Age | Sex | Survived | SexCode |
|----|------------------------------|--------|-----|--------|----------|---------|
| 12 | Aubert, Mrs Leontine Pauline | 1st | NaN | female | 1 | 1 |
| 13 | Barkworth, Mr Algernon H | 1st | NaN | male | 1 | 0 |

| | | | | | | |
|------|-------------------------|-----|-----|--------|-----|-----|
| 14 | Baumann, Mr John D | 1st | NaN | male | 0 | 0 |
| 29 | Borebank, Mr John James | 1st | NaN | male | 0 | 0 |
| 32 | Bradley, Mr George | 1st | NaN | male | 1 | 0 |
| ... | ... | ... | ... | ... | ... | ... |
| 1300 | Wiseman, Mr Phillippe | 3rd | NaN | male | 0 | 0 |
| 1302 | Yalsevac, Mr Ivan | 3rd | NaN | male | 1 | 0 |
| 1305 | Youssef, Mr Gerios | 3rd | NaN | male | 0 | 0 |
| 1306 | Zabour, Miss Hileni | 3rd | NaN | female | 0 | 1 |
| 1307 | Zabour, Miss Tamini | 3rd | NaN | female | 0 | 1 |

[557 rows x 6 columns]

```
[23]: import numpy as np
dataframe["Sex"]=dataframe["Sex"].replace('male',np.nan)
```

```
[24]: dataframe
```

```
[24]:
```

| | Name | PClass | Age | Sex | \ |
|------|---|--------|-------|--------|---|
| 0 | Allen, Miss Elisabeth Walton | 1st | 29.00 | female | |
| 1 | Allison, Miss Helen Loraine | 1st | 2.00 | female | |
| 2 | Allison, Mr Hudson Joshua Creighton | 1st | 30.00 | NaN | |
| 3 | Allison, Mrs Hudson JC (Bessie Waldo Daniels) | 1st | 25.00 | female | |
| 4 | Allison, Master Hudson Trevor | 1st | 0.92 | NaN | |
| ... | ... | ... | ... | ... | |
| 1308 | Zakarian, Mr Artun | 3rd | 27.00 | NaN | |
| 1309 | Zakarian, Mr Maprieder | 3rd | 26.00 | NaN | |
| 1310 | Zenni, Mr Philip | 3rd | 22.00 | NaN | |
| 1311 | Lievens, Mr Rene | 3rd | 24.00 | NaN | |
| 1312 | Zimmerman, Leo | 3rd | 29.00 | NaN | |

| | Survived | SexCode |
|------|----------|---------|
| 0 | 1 | 1 |
| 1 | 0 | 1 |
| 2 | 0 | 0 |
| 3 | 0 | 1 |
| 4 | 1 | 0 |
| ... | ... | ... |
| 1308 | 0 | 0 |
| 1309 | 0 | 0 |
| 1310 | 0 | 0 |
| 1311 | 0 | 0 |
| 1312 | 0 | 0 |

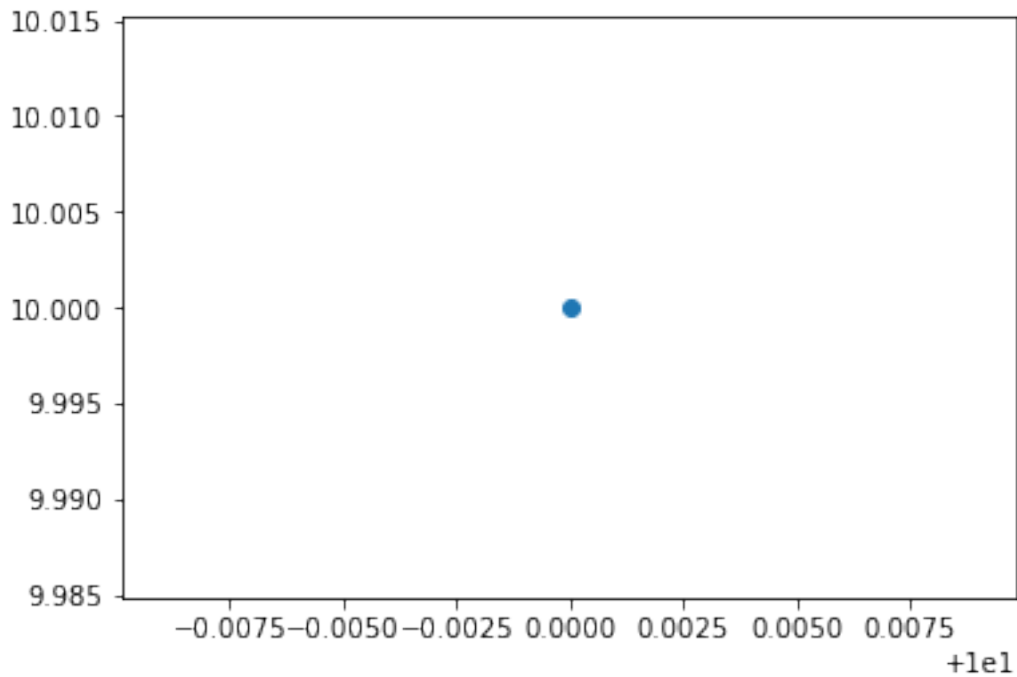
[1313 rows x 6 columns]

```
[25]: df.drop("Age",axis=1)
```

```
[25]: name
0 harshith
```

```
1    vamshi
2    Varun
```

```
[2]: import matplotlib.pyplot as plt
plt.scatter(10,10)
plt.show()
```



```
[5]: import pandas as pd
df=pd.DataFrame()
df["Name"]=["harshith", "bobby", "jitu", "sai", "teja"]
df["Age"]=[19,19,14,21,16]
def replace_name(val):
    if "Name"=="harshith":
        return 0
    else:
        return 1
df["Name"]=df["Name"].apply(replace_name,10)
df["Name"]
```

```
[5]: 0    1
1    1
2    1
3    1
4    1
Name: Name, dtype: int64
```

[: