

IPA (fundamental)

- GitHub link: [here](#)
- E-Mail: windkyle7@gmail.com

titanic

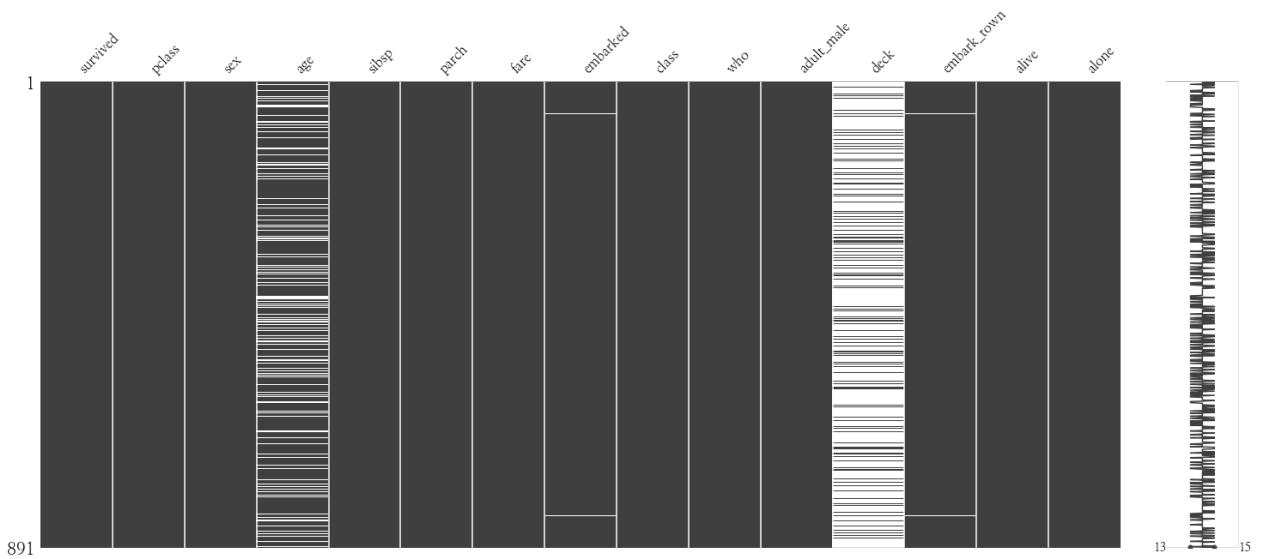
```
In [1]: import numpy as np
import pandas as pd
import seaborn as sns
import missingno as mno
import matplotlib.font_manager as fm
fm.rcParams['font.family'] = 'NanumMyeongjo'
```

```
In [2]: titanic = sns.load_dataset('titanic')
```

missing value

```
In [3]: %matplotlib inline
mno.matrix(titanic)
```

Out[3]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8a39944048>



titanic

```
In [4]: import pandas_profiling as pp
pp.ProfileReport(titanic)
```

Out[4]:

Overview

Dataset info

Number of variables	15
Number of observations	891
Total Missing (%)	6.5%
Total size in memory	80.6 KiB
Average record size in memory	92.6 B




Variables types

Numeric	5
Categorical	7
Boolean	3
Date	0
Text (Unique)	0
Rejected	0
Unsupported	0

Warnings

- `age` has 177 / 19.9% missing values Missing
- `deck` has 688 / 77.2% missing values Missing
- `fare` has 15 / 1.7% zeros Zeros
- `parch` has 678 / 76.1% zeros Zeros
- `sibsp` has 608 / 68.2% zeros Zeros
- Dataset has 107 duplicate rows Warning

Variables

<div>adult_male</div> <div>Boolean</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>2</div><div>0.2%</div><div>0.0%</div><div>0</div></div>	<div><div>Mean</div></div> <div><div>0.60269</div></div>	<div><div>True</div><div>False</div></div> <div><div>537</div><div>354</div></div>	<div></div> <div>Toggle details</div>
<div>age</div> <div>Numeric</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div><div>Infinite (%)</div><div>Infinite (n)</div></div> <div><div>89</div><div>10.0%</div><div>19.9%</div><div>177</div><div>0.0%</div><div>0</div></div>	<div><div>Mean</div><div>Minimum</div><div>Maximum</div><div>Zeros (%)</div></div> <div><div>29.699</div><div>0.42</div><div>80</div><div>0.0%</div></div>	<div></div> <div>Toggle details</div>	
<div>alive</div> <div>Categorical</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>2</div><div>0.2%</div><div>0.0%</div><div>0</div></div>	<div><div>no</div><div>yes</div></div> <div><div>549</div><div>342</div></div>	<div></div> <div>Toggle details</div>	
<div>alone</div> <div>Boolean</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>2</div><div>0.2%</div><div>0.0%</div><div>0</div></div>	<div><div>Mean</div></div> <div><div>0.60269</div></div>	<div><div>True</div><div>False</div></div> <div><div>537</div><div>354</div></div>	<div></div> <div>Toggle details</div>
<div>class</div> <div>Categorical</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>3</div><div>0.3%</div><div>0.0%</div><div>0</div></div>	<div><div>Third</div><div>First</div><div>Second</div></div> <div><div>491</div><div>216</div><div>184</div></div>	<div></div> <div>Toggle details</div>	
<div>deck</div> <div>Categorical</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>8</div><div>0.9%</div><div>77.2%</div><div>688</div></div>	<div><div>C</div><div>B</div><div>D</div><div>Other values (4)</div><div>(Missing)</div></div> <div><div>59</div><div>47</div><div>33</div><div>64</div><div>688</div></div>	<div></div> <div>Toggle details</div>	
<div>embark_town</div> <div>Categorical</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>4</div><div>0.4%</div><div>0.2%</div><div>2</div></div>	<div><div>Southampton</div><div>Cherbourg</div><div>Queenstown</div><div>(Missing)</div></div> <div><div>644</div><div>168</div><div>77</div><div>2</div></div>	<div></div> <div>Toggle details</div>	
<div>embarked</div> <div>Categorical</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div></div> <div><div>4</div><div>0.4%</div><div>0.2%</div><div>2</div></div>	<div><div>S</div><div>C</div><div>Q</div><div>(Missing)</div></div> <div><div>644</div><div>168</div><div>77</div><div>2</div></div>	<div></div> <div>Toggle details</div>	
<div>fare</div> <div>Numeric</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div><div>Infinite (%)</div><div>Infinite (n)</div></div> <div><div>248</div><div>27.8%</div><div>0.0%</div><div>0</div><div>0.0%</div><div>0</div></div>	<div><div>Mean</div><div>Minimum</div><div>Maximum</div><div>Zeros (%)</div></div> <div><div>32.204</div><div>0</div><div>512.33</div><div>1.7%</div></div>	<div></div> <div>Toggle details</div>	
<div>parch</div> <div>Numeric</div>	<div><div>Distinct count</div><div>Unique (%)</div><div>Missing (%)</div><div>Missing (n)</div><div>Infinite (%)</div><div>Infinite (n)</div></div> <div><div>7</div><div>0.8%</div><div>0.0%</div><div>0</div><div>0.0%</div><div>0</div></div>	<div><div>Mean</div><div>Minimum</div><div>Maximum</div><div>Zeros (%)</div></div> <div><div>0.38159</div><div>0</div><div>6</div><div>76.1%</div></div>	<div></div> <div>Toggle details</div>	


```
-----
dtypes: bool(2), category(2), float64(2), int64(4), object(5)
memory usage: 80.6+ KB
```

```
sex  class  0  1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23
```

```
In [6]: group = titanic.groupby(['sex','class']).survived
total = group.sum()
```

```
In [7]: total
```

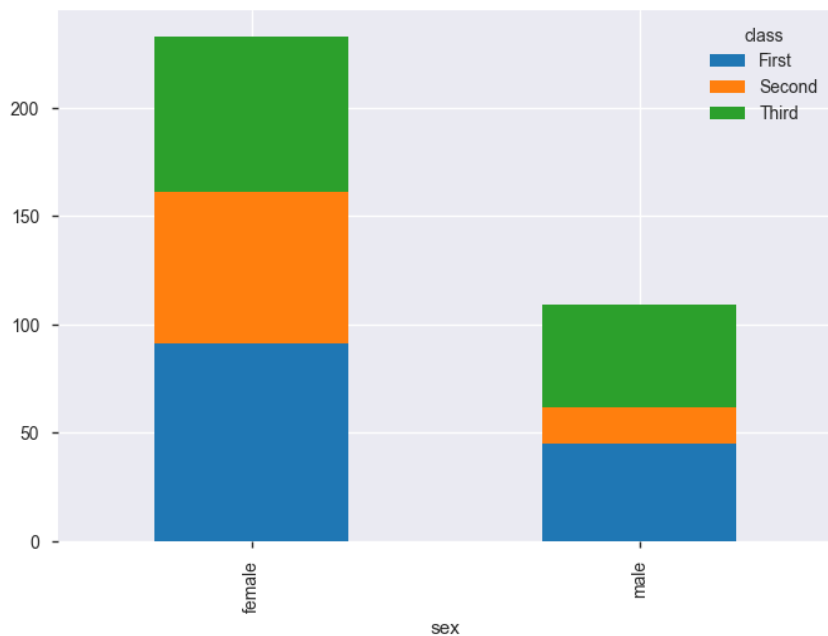
```
Out[7]: sex      class
female  First     91
        Second    70
        Third     72
male    First     45
        Second    17
        Third     47
Name: survived, dtype: int64
```

```
In [8]: total.unstack()
```

```
Out[8]:
      class  First  Second  Third
sex
female    91     70     72
male      45     17     47
```

```
In [9]: total.unstack().plot.bar(stacked=True)
```

```
Out[9]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8a5fe0ef98>
```



```
In [10]: titanic.embarked.value_counts()
```

```
Out[10]: S      645
C      169
Q       77
Name: embarked, dtype: int64
```

```
In [11]: titanic.embarked.map({'S':0,'C':1,'Q':2})
```

```
Out[11]: 0      0
1      1
2      0
3      0
4      0
5      2
6      0
7      0
8      0
9      1
10     0
11     0
12     0
13     0
14     0
15     0
16     2
17     0
18     0
19     1
20     0
21     0
22     2
23     0
```

```

23 0
24 0
25 0
26 1
27 0
28 2
29 0
..
861 0
862 0
863 0
864 0
865 0
866 1
867 0
868 0
869 0
870 0
871 0
872 0
873 0
874 1
875 1
876 0
877 0
878 0
879 1
880 0
881 0
882 0
883 0
884 0
885 2
886 0
887 0
888 0
889 1
890 2
Name: embarked, Length: 891, dtype: int64

```

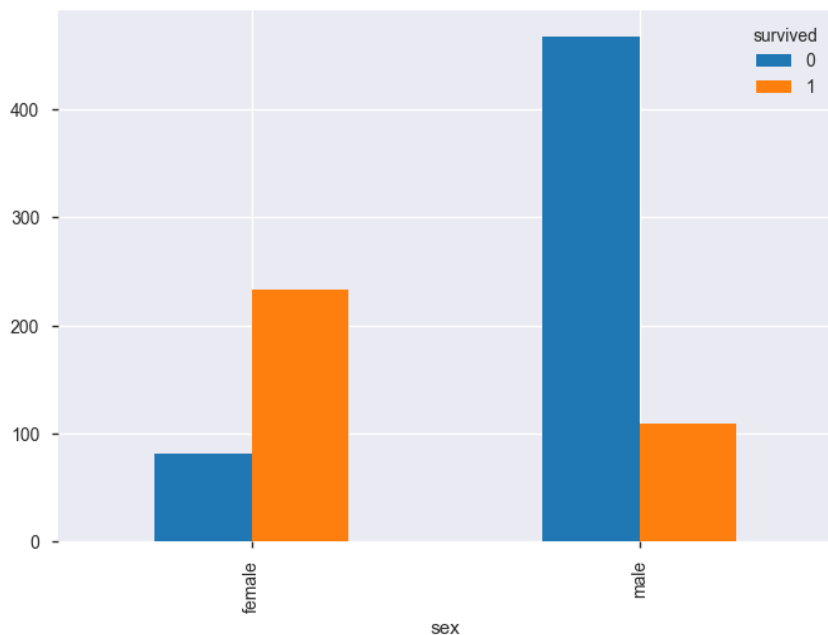
```
In [12]: titanic.pivot_table('survived', 'sex', aggfunc=sum)
```

```
Out[12]:
```

survived	
sex	
female	233
male	109

```
In [13]: table = pd.crosstab(titanic.sex, titanic.survived)
table.plot.bar()
```

```
Out[13]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8a5fe0e630>
```



```
In [14]: table.stack()
```

```
Out[14]: sex    survived
female  0         81
         1        233
male    0        468
         1        109
dtype: int64
```

```
In [15]: table.unstack()
```

```
Out[15]: survived sex
0         female    81
```

```
male 468
1 female 233
male 109
dtype: int64
```

```
In [16]: survived_group = titanic.groupby(['sex', 'survived'])
total = survived_group.sum().unstack()
total
```

Out[16]:

	pclass		age		sibsp		parch		fare		adult_male		alone	
survived	0	1	0	1	0	1	0	1	0	1	0	1	0	1
sex														
female	231	447	2112.00	6886.42	98	120	84	120	1864.9752	12101.6876	0.0	0.0	27.0	99.0
male	1159	220	14637.83	2978.42	206	42	97	39	10277.7447	4449.5418	449.0	88.0	347.0	64.0

```
In [17]: total.stack()
```

Out[17]:

		pclass	age	sibsp	parch	fare	adult_male	alone
sex	survived							
female	0	231	2112.00	98	84	1864.9752	0.0	27.0
	1	447	6886.42	120	120	12101.6876	0.0	99.0
male	0	1159	14637.83	206	97	10277.7447	449.0	347.0
	1	220	2978.42	42	39	4449.5418	88.0	64.0

```
In [18]: total.plot.bar(stacked=True)
```

Out[18]: <matplotlib.axes._subplots.AxesSubplot at 0x7f8a5fcc9390>

