```
from google.co lab import files up
load = f i les.uploadO
```

```
| 파일 선택 | 파일 2 개
• test.csv(text/csv) - 28629 bytes, last modified: 2025. 4. 9. - 100% done
• train.csv(text/csv) - 61194 bytes, last modified: 2025. 4. 9. - 100% done Saving
test.csv to test.csv
Saving train.csv to train.csv
```

```
import pandas as pd
import numpy as np

t = pd.read_csv(1 train.csv') te=
pd.read_csv(1 test.csv1)
```

```
t.headO
```

	Passenger Id	Surv i ved	Pc l ass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cab i	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0			○ :	7.2500	Na	
1	2	1	1	Cumings, Mrs. John Bradley (Florence	female	38.0		0	PC 17599	71.2833	C8	

Next steps: ◦ **View recommended plots**

```
t.shape
```

```
(891 , 12)
```

```
te.shape
```

```
(418, 11)
```

```
t.info()
```

```
<cl ass 1 pandas.c 아 e.frame.DataFrame1 〉 Range Index : 891 entries, 0 to 890 Data columns (total 12 columns):
#   Co lumn          Non-Nu11 I Count   Dtype
---  ------          ----------------   -----
0   Passenger Id     891 non nul I      int64
1   Survived         891 non null       int64
2   Pc l ass         891 non nul I      int64
3   Name             891 non null       object
4   Sex              891 non nul I      object
5   Age              714 non null       fl oat64
6   SibSp            891 non nul I      int64
7   Parch            891 non nul I      int64
8   Ticket           891 non null       object
9   Fare             891 non nul I      fl oat64
10  Cabin            204 non null       object
11  Embarked         889 non nul I      object
dtypes : float64(2), int64(5), object(5) memory usage : 83.7+ KB
```

```
te.info()
```

```
<cl ass 1 pandas.c 아 e.f rame.DataFrame'> Range Index : 418 entries, 0 to 417
Data columns (total 11 columns):
#   Co lumn          Non-Nu11     Count   Dtype
---  ------          ----------   -----   -----
0   Passenger Id     418 non      nul I   int64
1   Pc l ass         418 non      null    int64
2   Name             418 non      nul I   object
3   Sex              418 non      null    object
4   Age              332 non      nul I   fl oat64
5   SibSp            418 non      nul I   int64
6   Parch            418 non      nul I   int64
7   Ticket           418 non      nul I   object
8   Fare             417 non      null    fl oat64
9   Cabin            91 non-null          object
10  Embarked         418 non-nu11         object
dtypes : float64(2), int64(4), object(5) memory usage : 36.0+ KB
```

```
t.descr ibe()
```

	Passenger Id	Survived	Pc l ass	Age	SibSp	Parch	Fare 西
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000000 H
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204208
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693429
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910400
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454200
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329200

te.descr ibe()

	Passenger Id	Pci ass	Age	SibSp	Parch	Fare 西
count	418.000000	418.000000	332.000000	418.000000	418.000000	417.000000 m
mean	1100.500000	2.265550	30.272590	0.447368	0.392344	35.627188
std	120.810458	0.841838	14.181209	0,896760	0.981429	55.907576
min	892.000000	1.000000	0.170000	0.000000	0.000000	0.000000
25%	996.250000	1.000000	21.000000	0.000000	0.000000	7.895800
50%	1100.500000	3.000000	27.000000	0.000000	0.000000	14.454200
75%	1204.750000	3.000000	39.000000	1.000000	0.000000	31.500000
max	1309.000000	3.000000	76.000000	8.000000	9.000000	512.329200

t. isnul IO.sumO

Passenger Id	0
Survived	0
Pc l ass	0
Name	0
Sex	0
Age	177
SibSp	0
Parch	0
Ticket	0
Fare	0
Cabin	687
Embarked	2
dtype： int64	

te.isnull().sum()

Passenger Id	0
Pc l ass	0
Name	0
Sex	0
Age	86
SibSp	0
Parch	0
Ticket	0
Fare	1
Cabin	327
Embarked	0
dtype： int64	

for col in t.columns：

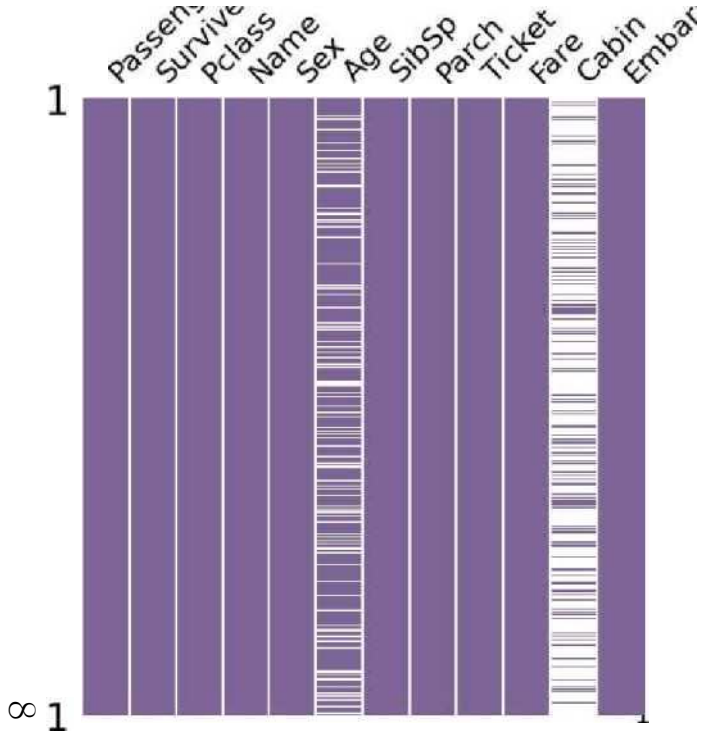
msg ='column： {:>15} {:.2f}%'.format(col, 100*(t[col].isnull().sum()/t[col].shape[0])) print(msg)

co lumn:	Passenger Id 0.00%
co lumn:	Survived 0.00%
co lumn:	Pc l ass 0.00%
co lumn:	Name 0.00%
co lumn:	Sex 0.00%
co lumn:	Age 19.87%
co lumn:	SibSp 0.00%
co lumn:	Parch 0.00%
co lumn:	Ticket 0.00%
co lumn:	Fare 0.00%
co lumn:	Cabin 77.10%
co lumn:	Embarked 0.22%

import missingno as msno

msno.matrix(df=t.iloc[:,:] , figsize=(6,6),cok)r=(0.5,0.4,0.6))

<Axes : >



t[[1 Pc I ass1,'Survived1]].groupbyi['Pci ass1J，as_index=True).count()

Pc I ass	Survived 떤 Q
1	216
2	184
3	491

t[[1 Pc I ass1,'Survived1]].groupby(['Pci ass1],as_ i ndex=Tr ue).sum()

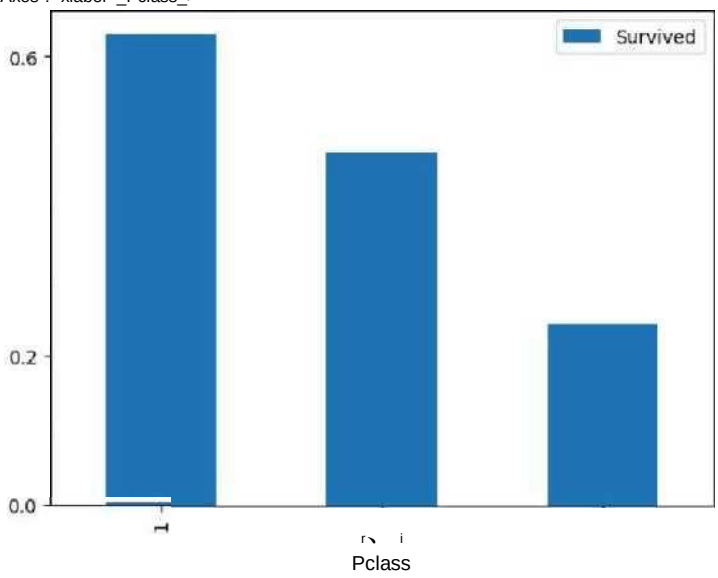
Pc I ass	Survived 떤 Q
1	**136**
2	87
3	119

pd.crosstab(t['Pclass'],t['Survived1],margins=True).style.backQround_gradient(cmap='summer_r1)

Survived 0		1 Al I
Pc I ass		
1	80	136 216
2	97	87 184
3	372	119 491
Ml	549	342 891

t[['Pclass','Survived']].groupby(['Pci ass'],as_index=True).mean().plot.bar()

<Axes : xlabel=_Pclass_>



import seaborn as sns sns.countplotCPclass',hue='Survived'data=t) t[[1 Sex1, 'Survived1]].groupbyd'Sex1],as_index=True).mean()

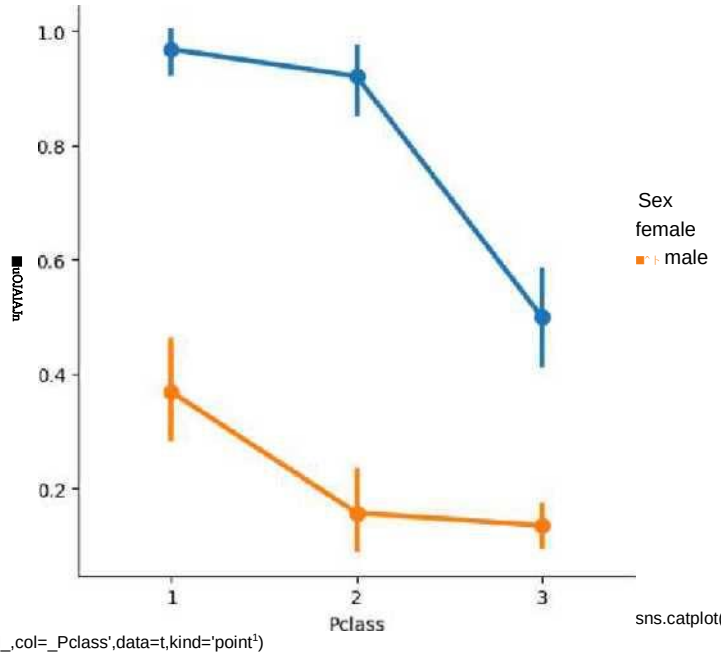
	Survived Q
Sex	|0
female	**0.742038**
male	**0.188908**

pd.crosstab(t['Sex'],t['SurvIved'],maroins=True).style.background_gradient(cmap='summer_r _)

Survived	0 1		Al I
Sex			
female	**81**	233	**314**
male	**468**	**109**	**577**
All	**549**	342	**891**

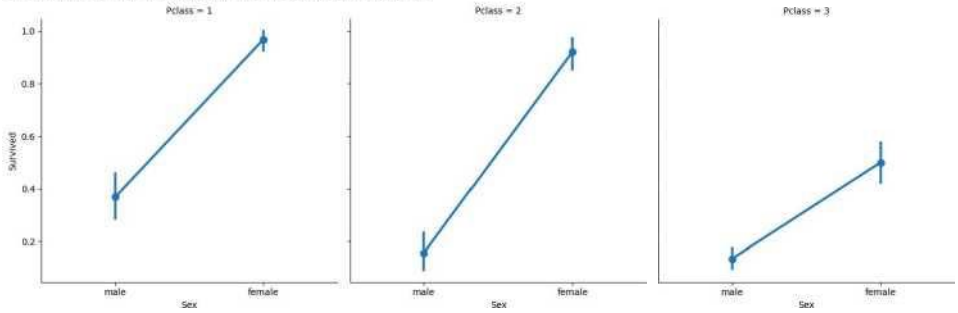
import seaborn as sns
sns.catp lot (x~ Pci ass' ^'Survived1 ,hue='Sex' ,data=t,kind=_ point')

<seab ○├ n.axisgrid.FacetGrid at 0x7b2017e12d70>



sns.catplot(x='Sex',y='Su
「vived_,col=_Pclass',data=t,kind='point1)

```
<seaborn.axisgrid.FacetGrid at 0x7b2017ce25c0>
```

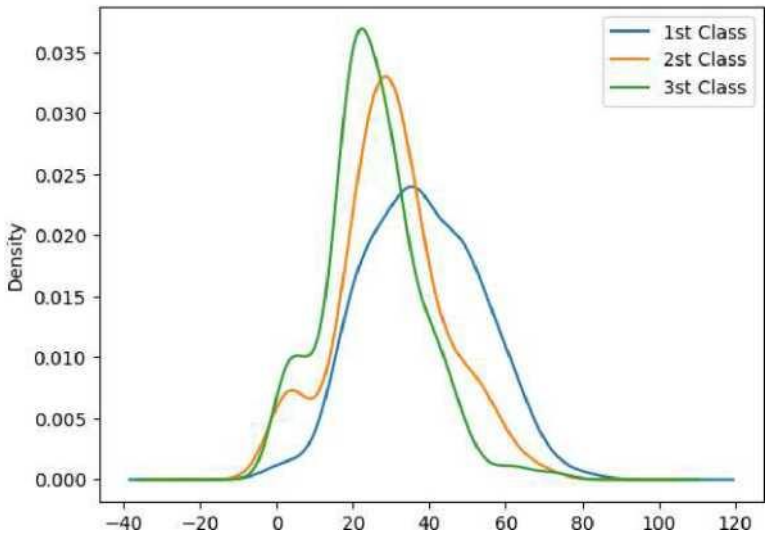



```
print(t['Age'].max())
print(t['Age'].min())
t['Age'].mean()
```

```
80.0
0.42
29.69911764705882
```

```
import matplotlib.pyplot as pit
t[ 'Age¹ ] [t[ 'Pci ass' ]=1] .plot (kind='kde') t [ 'Age' ]
[t[  'Pclass¹  ]=2]  .plot(kind='kde¹)  t  [  'Age¹  ]  [t
[ 'Pciass' ]=3] .plot(kind='kde')
pit.legend([¹11st Class','2st Class*,'3st Class¹,])
```

```
<matplotlib.legend.Legend at 0x7b2017b0d150>
```



코드 표시