```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
df = pd.read_csv("test.csv")
df.head()
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

Next steps: Generate code with df New interactive sheet

```
df.isnull().sum()
```

	0
PassengerId	0
Pclass	0
Name	0
Sex	0
Age	86
SibSp	0
Parch	0
Ticket	0
Fare	1
Cabin	327
Embarked	0

dtype: int64

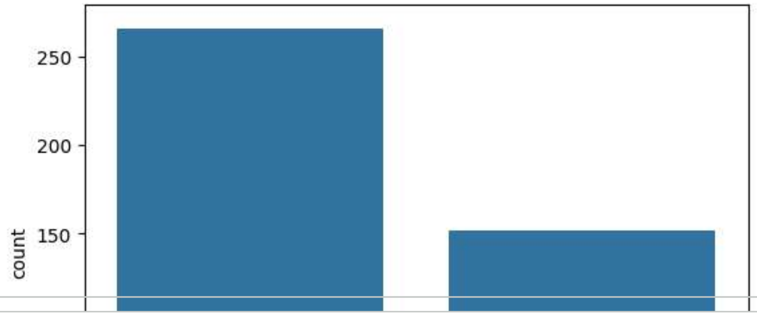
```
df.drop_duplicates
```

```
pandas.core.frame.DataFrame.drop_duplicates
def drop_duplicates(subset: Hashable | Sequence[Hashable] | None=None, *, keep: DropKeep='first',
inplace: bool=False, ignore_index: bool=False) -> DataFrame | None
```

```
/usr/local/lib/python3.12/dist-packages/pandas/core/frame.py
Return DataFrame with duplicate rows removed.

Considering certain columns is optional. Indexes, including time indexes
are ignored.
```

```
sns.countplot(x='Sex',data=df)
plt.show()
```

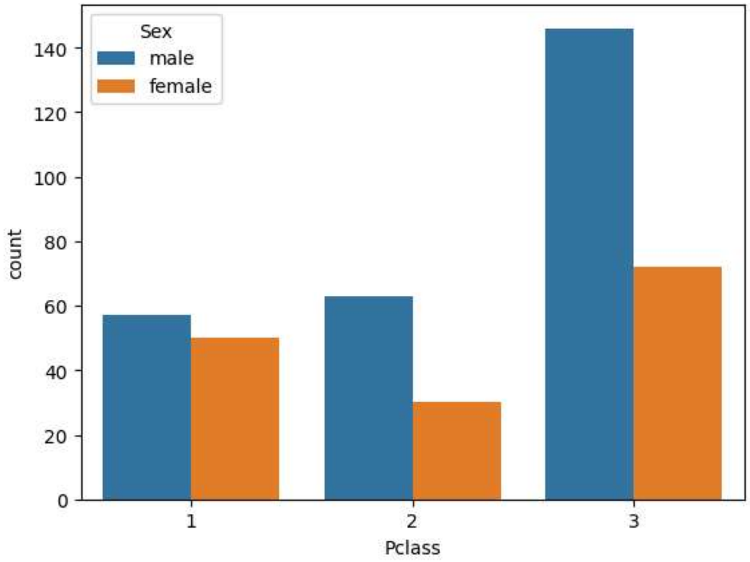


```
sns.countplot(x='Pclass', hue='Sex', data=df)
plt.title=('Pclass: male vs female')
plt.show
```

```
matplotlib.pyplot.show
def show(*args, **kwargs) -> None
```

```
/usr/local/lib/python3.12/dist-packages/matplotlib/pyplot.py
Display all open figures.

Parameters
----------
block : bool, optional
```



```
print('Oldest person Sex was of:',df['Age'].max())
print('Youngest person Sex was of:',df['Age'].min())
print('Average person Sex was of:',df['Age'].mean())
```

```
Oldest person Sex was of: 76.0
Youngest person Sex was of: 0.17
Average person Sex was of: 30.272590361445783
```

```
df['Initial']=0
for i in df:
    df['Initial']=df.Name.str.extract('([A-Za-z]+)\.')
```

```
<>:3: SyntaxWarning: invalid escape sequence '\.'
<>:3: SyntaxWarning: invalid escape sequence '\.'
/tmp/ipython-input-3730373830.py:3: SyntaxWarning: invalid escape sequence '\.'
  df['Initial']=df.Name.str.extract('([A-Za-z]+)\.')
```

```
pd.crosstab(df.Initial,df.Sex).T.style.background_gradient(cmap='summer_r')
```

Initial	Col	Dona	Dr	Master	Miss	Mr	Mrs	Ms	Rev
Sex									
female	0	1	0	0	78	0	72	1	0
male	2	0	1	21	0	240	0	0	2

```
df['Initial'].replace(['Mlle','Mme','Ms','Dr','Major','Lady','Countess',
                               'Jonkheer','Col','Rev','Capt','Sir','Don'],['Miss',
                                'Miss','Miss','Mr','Mr','Mrs','Mrs','Other','Other','Other','Mr','Mr','Mr'],inplace=True)
```

```
df.groupby('Initial')['Age'].mean()
```

	Age
Initial	
Dona	39.000000
Master	7.406471
Miss	21.774844
Mr	32.114130
Mrs	38.903226
Other	42.750000

dtype: float64

```
df.loc[(df.Age.isnull()) & (df.Initial=='Mr'),'Age']=33
df.loc[(df.Age.isnull()) & (df.Initial=='Mrs'),'Age']=36
df.loc[(df.Age.isnull()) & (df.Initial=='Master'),'Age']=5
df.loc[(df.Age.isnull()) & (df.Initial=='Miss'),'Age']=22
df.loc[(df.Age.isnull()) & (df.Initial=='Other'),'Age']=46
```

```
df.Age.isnull().any()
```

```
np.False_
```

```
pd.crosstab(df.SibSp,df.Pclass).style.background_gradient('summer_r')
```

Pclass	1	2	3
SibSp			
0	61	62	160
1	42	27	41
2	3	4	7
3	1	0	3
4	0	0	4
5	0	0	1
8	0	0	2