1- Preprocessing Part:

```python
import pandas as pd
df=pd.read_excel("titanic-passengers.xlsx")
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	343	No	2	Collander, Mr. Erik Gustaf	male	28.0	0	0	248740	13.00
1	76	No	3	Moen, Mr. Sigurd Hansen	male	25.0	0	0	348123	7.65
2	641	No	3	Jensen, Mr. Hans Peder	male	20.0	0	0	350050	7.85
3	568	No	3	Palsson, Mrs. Nils (Alma	female	29.0	0	4	349909	21.07

```python
df.columns
```

```
Index(['PassengerId', 'Survived', 'Pclass', 'Name', 'Sex', 'Age', 'SibSp',
       'Parch', 'Ticket', 'Fare', 'Cabin', 'Embarked'],
```

Enregistrement...   ✕

```python
print(df.isnull().sum())
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

```python
print(df.isnull().sum().sum())
```

```
866
```

```python
print(df['Age'].isnull().sum())
```

```
177
```

```python
print(df['Cabin'].isnull().sum())
```

```
687
```

```python
df["Age"].fillna(df["Age"].mean(),inplace=True)
df.head(20)
```

Enregistrement...   ×

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Tick
0	343	No	2	Collander, Mr. Erik Gustaf	male	28.000000	0	0	2487
1	76	No	3	Moen, Mr. Sigurd Hansen	male	25.000000	0	0	3481
2	641	No	3	Jensen, Mr. Hans Peder	male	20.000000	0	0	3500
3	568	No	3	Palsson, Mrs. Nils (Alma Cornelia Berglund)	female	29.000000	0	4	3499
4	672	No	1	Davidson, Mr. Thornton	male	31.000000	1	0	F. 127
5	105	No	3	Gustafsson, Mr. Anders Vilhelm	male	37.000000	2	0	31012

```
df['Embarked'].value_counts()
```

```
S    644
C    168
Q     77
```

Lovell, Mr.

```
df['Cabin'].value_counts()
```

```
B96 B98       4
C23 C25 C27   4
G6            4
F33           3
F2            3
             ..
C111          1
C50           1
D49           1
A34           1
E12           1
Name: Cabin, Length: 147, dtype: int64
```

```
df['Cabin'].fillna('G6',inplace=True)
df['Cabin'].value_counts()
```

```
G6           691
B96 B98        4
C23 C25 C27    4
F33            3
F2             3
             ...
C111           1
```

```
        C50                1
        D49                1
        A34                1
        E12                1
        Name: Cabin, Length: 147, dtype: int64
```

```
df['Embarked'].fillna('S',inplace=True)
df['Embarked'].value_counts()
```

```
        S    646
        C    168
        Q     77
        Name: Embarked, dtype: int64
```
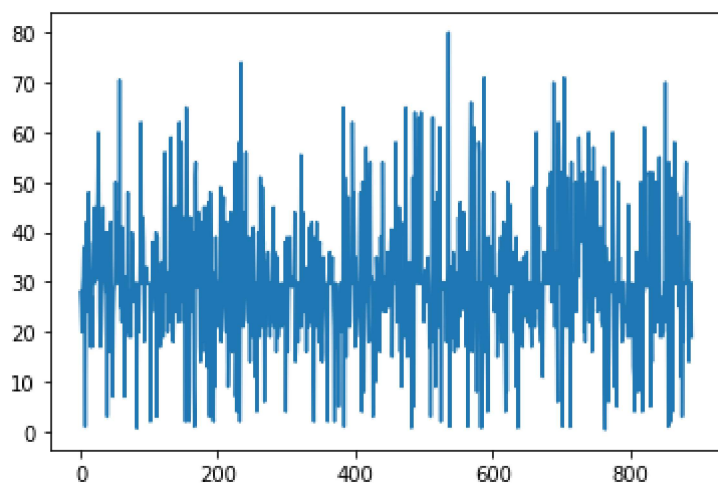
```
df.isnull().sum()
```

```
        PassengerId    0
        Survived       0
        Pclass         0
        Name           0
        Sex            0
        Age            0
        SibSp          0
        Parch          0
        Ticket         0
        Fare           0
        Cabin          0
        Embarked       0
        dtype: int64
```

Enregistrement...   ✕

d

2- Visualization Part:

```
import matplotlib.pyplot as plt
plt.plot(df['Age'])
plt.show()
```



```
import matplotlib.pyplot as plt
```

```
v=df['Sex'].value_counts()
v.plot.bar(rot=45)
```
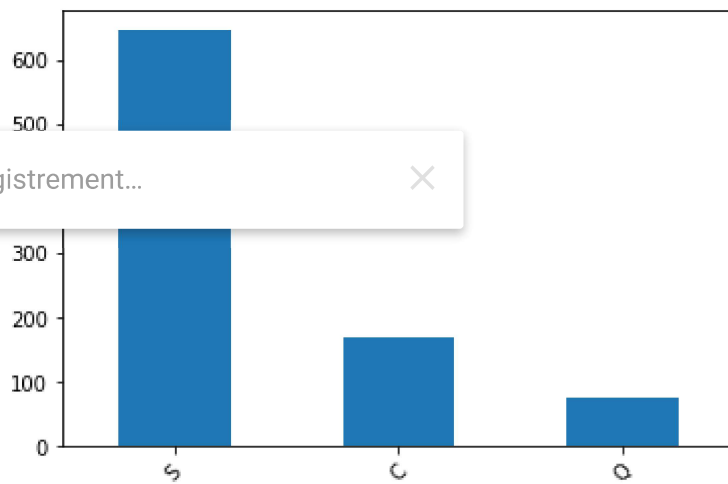
```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe857519f10>
```
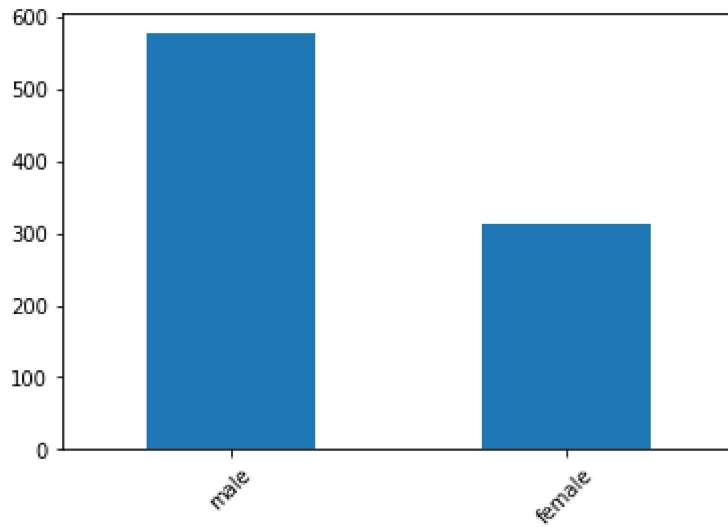


```
import matplotlib.pyplot as plt
v=df['Embarked'].value_counts()
v.plot.bar(rot=45)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe857556310>
```
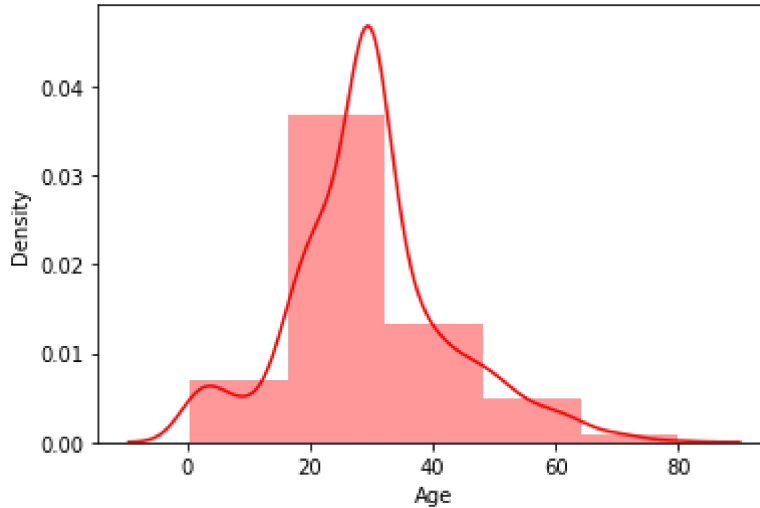


```
import matplotlib.pyplot as plt
df['Age'].plot.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7fe857477290>
```



```
import seaborn as sns
sns.distplot(df['Age'],bins=5,hist=True,kde=True,color='red')
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2557: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7fe8499ba050>
```
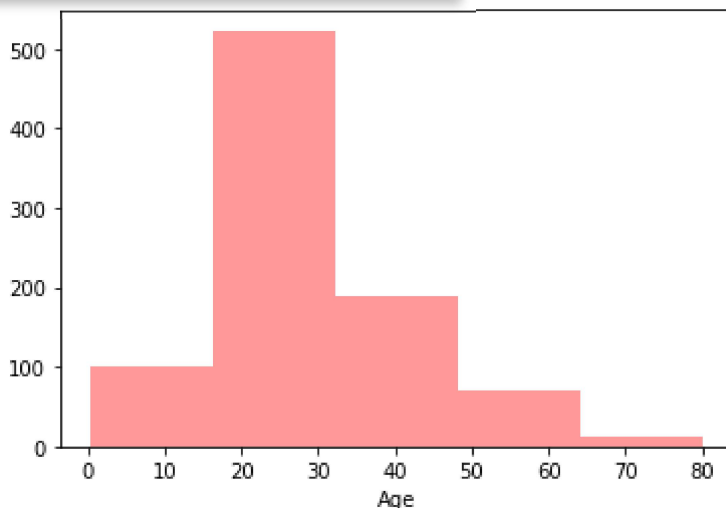


```
import seaborn as sns
sns.distplot(df['Age'],bins=5,hist=True,kde=False,color='red')
```
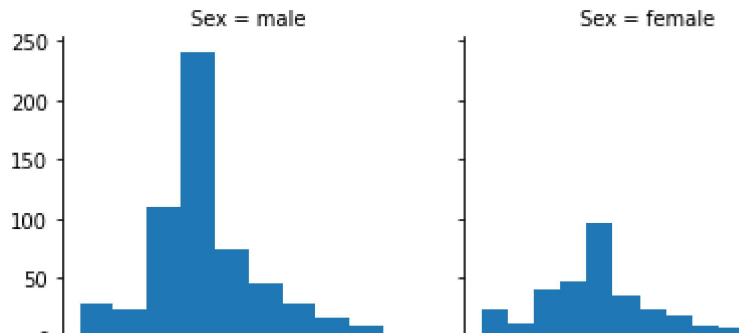
```
                              packages/seaborn/distributions.py:2557: FutureWarning:
                              ning)
                              sSubplot at 0x7fe848f7bd90>
```

Enregistrement... ✕


```
import matplotlib.pyplot as plt
import seaborn as sns
grid=sns.FacetGrid(df,col='Sex')
grid.map(plt.hist,'Age')
```

```
<seaborn.axisgrid.FacetGrid at 0x7fe8466c20d0>
```
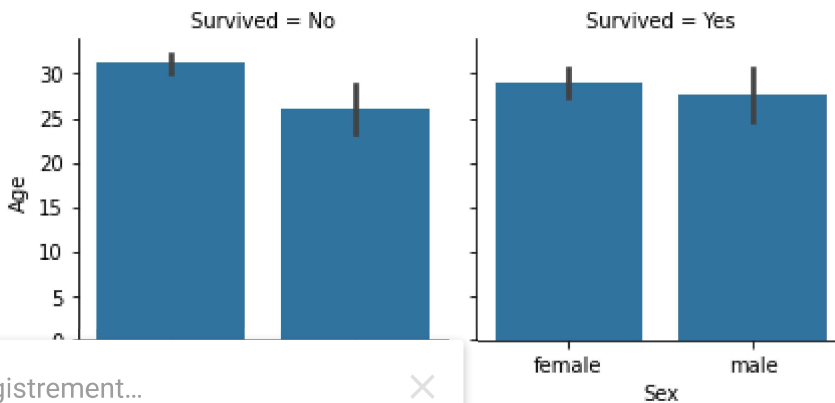


```
import matplotlib.pyplot as plt
import seaborn as sns
grid=sns.FacetGrid(df,col='Survived')
grid.map(sns.barplot,'Sex','Age')
grid.add_legend()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:643: UserWarning: Using t
    warnings.warn(warning)
<seaborn.axisgrid.FacetGrid at 0x7fe8466d0f90>
```
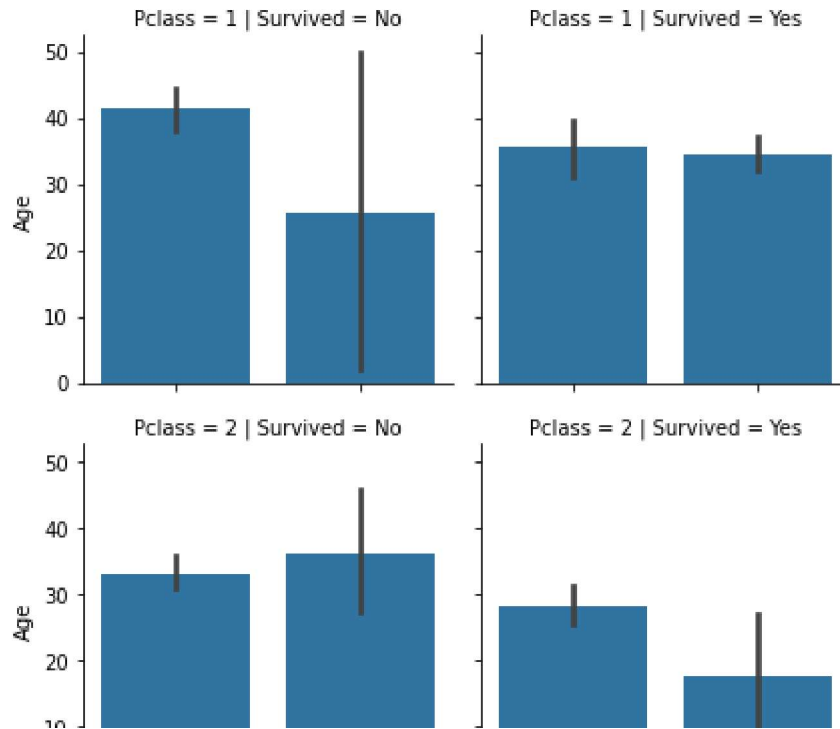

Enregistrement... ✕

```
import matplotlib.pyplot as plt
import seaborn as sns
grid=sns.FacetGrid(df,row='Pclass',col='Survived')
grid.map(sns.barplot,'Sex','Age')
grid.add_legend()
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/axisgrid.py:643: UserWarning: Using t
  warnings.warn(warning)
<seaborn.axisgrid.FacetGrid at 0x7fe83dc78b10>
```
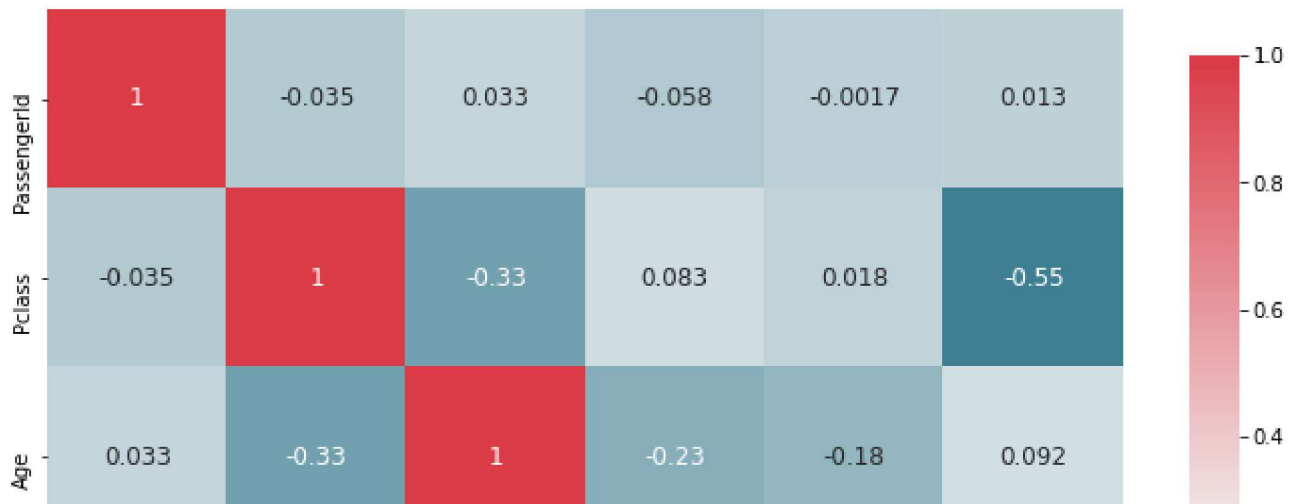


```python
def plot_correlation_map(df):
  corr=df.corr()
  s,ax=plt.subplots(figsize=(12,10))
  cmap=sns.diverging_palette(220,10,as_cmap=True)
  s=sns.heatmap(
```
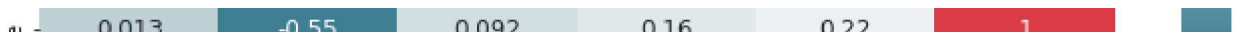
Enregistrement…   ×

```python
      cbar_kws={'shrink':.9},
      ax=ax,
      annot=True,
      annot_kws={'fontsize':12}
  )
plot_correlation_map(df)
```

Interprétation: cette fonction montre la corrélation entre les variables de la data . La corrélation varie de -1 à +1. Les valeurs les plus proches de 0 signifient qu'il n'y a pas une relation dans le changement entre les deux variables c-à-d une correlation neutre . Alors que les valeurs les plus proches de 1, les deux variables changent dans le meme sens c'est une correlation positive positivement. On trouve aussi des valeurs qui sont proche de -1, ici les deux variables changent dans des directions opposées c'est une corrélation négative. mais au lieu des deux, une variable augmentera à mesure que l'autre augmentera.la correlation entre chaque variable avec elle-même est une corrélation parfaite(Les diagonales sont à 1).


```
df[["Survived","Pclass"]].groupby(["Survived"], as_index=False).mean()
```

Enregistrement...   ✕

0	No	2.531876
1	Yes	1.950292

```
df["Surname"] = df["Name"].str.split(".").str.get(0)
df["Surname"]
df["title"] = df["Surname"].str.split(",").str.get(-1)
df["title"]
```

```
0        Mr
1        Mr
2        Mr
3       Mrs
4        Mr
       ...
886     Mrs
887      Mr
888    Miss
889      Mr
890    Miss
Name: title, Length: 891, dtype: object
```

```
df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fa
0	343	No	2	Collander, Mr. Erik Gustaf	male	28.0	0	0	248740	13.00
1	76	No	3	Moen, Mr. Sigurd Hansen	male	25.0	0	0	348123	7.65
2	641	No	3	Jensen, Mr. Hans Peder	male	20.0	0	0	350050	7.85
3	568	No	3	Palsson, Mrs. Nils (Alma Cornelia Berglund)	female	29.0	0	4	349909	21.07
4	672	No	1	Davidson, Mr. Thornton	male	31.0	1	0	F.C. 12750	52.00

```python
df=df.drop('Name',axis=1)
df=df.drop('Surname',axis=1)
df.head()
```
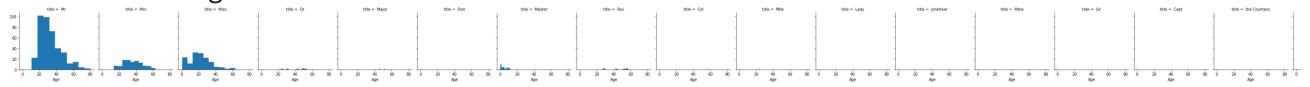
	PassengerId	Survived	Pclass	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin
			2	male	28.0	0	0	248740	13.0000	NaN
			3	male	25.0	0	0	348123	7.6500	F G73
2	641	No	3	male	20.0	0	0	350050	7.8542	NaN
3	568	No	3	female	29.0	0	4	349909	21.0750	NaN
4	672	No	1	male	31.0	1	0	F.C. 12750	52.0000	B71

Enregistrement...

```python
import seaborn as sns
import matplotlib.pyplot as plt
g=sns.FacetGrid(df,col='title')
g.map(plt.hist,"Sex")
```

```
<seaborn.axisgrid.FacetGrid at 0x7f0951e654d0>
```



```python
import seaborn as sns
import matplotlib.pyplot as plt
g=sns.FacetGrid(df,col='title')
g.map(plt.hist,"Age")
```

```
<seaborn.axisgrid.FacetGrid at 0x7f09515d17d0>
```


Enregistrement...