```
import pandas as pd
df=pd.read_csv("/content/train (1).csv",header="infer")
print(df.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
None
```

```
#Ignore Tuples
df1=df.copy()

df1=df.dropna(axis=0)
print(df1.shape,df.shape)
df.info()
```

```
(183, 12) (891, 12)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          714 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        204 non-null    object 
 11  Embarked     889 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
#Assign a Global Constant to missing values
df2=df.copy()

df2.loc[:,"Embarked"]
df2.loc[df2.loc[:,"Age"].isna(),"Age"]=31
df2.loc[df2.loc[:,"Embarked"].isna(),"Embarked"]='S'

c=['C85','C123','C103','G6','A6','D56']
import random
i=random.randint(0,5)
df2.loc[df2.loc[:,"Cabin"].isna(),"Cabin"]=c[i]

df2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        891 non-null    object 
 11  Embarked     891 non-null    object 
```

```
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
#Assign mean or mode of the data to missing places
df3=df.copy()
import numpy as np

mean_age=np.mean(df3.loc[~df3.loc[:,"Age"].isna(),"Age"].values)
print(mean_age)

df3.loc[df3.loc[:,"Age"].isna(),"Age"]=mean_age

mode_cab=df3.loc[:,"Cabin"].mode()[0]
print(mode_cab)

df3.loc[df3.loc[:,"Cabin"].isna(),"Cabin"]=mode_cab

mode_emb=df3.loc[:,"Embarked"].mode()[0]
print(mode_emb)

df3.loc[df3.loc[:,"Embarked"].isna(),"Embarked"]=mode_emb

df3.info()
```

```
29.69911764705882
B96 B98
S
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        891 non-null    object 
 11  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

```
df4=df.copy()

mean_age1=np.mean(df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==0),"Age"].values)
print(mean_age1)

df4.loc[(df4.loc[:,"Age"].isna()) & (df4["Survived"]==0),"Age"]=mean_age1

mean_age2=np.mean(df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==1),"Age"].values)
print(mean_age2)

df4.loc[(df4.loc[:,"Age"].isna()) & (df4["Survived"]==1),"Age"]=mean_age2

mode_cab1=df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==0),"Cabin"].mode()[0]
print(mode_cab1)
df4.loc[(df4.loc[:,"Cabin"].isna()) & (df4["Survived"]==0),"Cabin"]=mode_cab1

mode_cab2=df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==1),"Cabin"].mode()[0]
print(mode_cab2)
df4.loc[(df4.loc[:,"Cabin"].isna()) & (df4["Survived"]==1),"Cabin"]=mode_cab2

mode_emb1=df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==0),"Embarked"].mode()[0]
print(mode_emb1)
df4.loc[(df4.loc[:,"Embarked"].isna()) & (df4["Survived"]==0),"Embarked"]=mode_emb1

mode_emb2=df4.loc[(~df4.loc[:,"Age"].isna()) & (df4["Survived"]==1),"Embarked"].mode()[0]
print(mode_emb2)
df4.loc[(df4.loc[:,"Embarked"].isna()) & (df4["Survived"]==1),"Embarked"]=mode_emb2

df4.info()
```

```
30.62617924528302
28.343689655172415
C124
B96 B98
S
S
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object 
 4   Sex          891 non-null    object 
 5   Age          891 non-null    float64
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   ticket       891 non-null    object 
 9   Fare         891 non-null    float64
 10  Cabin        891 non-null    object 
 11  Embarked     891 non-null    object 
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

✓ 0s completed at 8:14 PM