FOR THE TITANIC DATASET PERFORM THE FOLLOWING

In [10]: # Package imports import pandas as pd #import matplotlib.pyplot as plt import missingno as msno %matplotlib inline #Importing the required dataset In [5]: titanic_df = pd.read_csv("titanic.csv") titanic_df Out[5]: survived sibsp fare embarked class who adult male pclass age parch sex 0 0 7.2500 S Third 3 male 22.0 0 man Tru€ 1 1 38.0 71.2833 C female 0 First Fals€ woman 2 1 female 26.0 0 7.9250 S Third woman Fals€ 3 1 female 35.0 0 53.1000 S False First woman 4 0 3 male 35.0 0 0 8.0500 S Third man Tru€ 886 0 2 27.0 13.0000 S Second male 0 man Tru€ 887 female 19.0 0 30.0000 S False First woman S 0 888 3 NaN 1 23.4500 female Third woman Fals€ 889 30.0000 1 male 26.0 0 First True man 890 0 3 male 32.0 0 0 7.7500 Q Third man Tru€ 891 rows × 15 columns

1.Display the number of missing values for each feature in the dataset

In [6]: titanic_df.info()

<class 'pandas.core.frame.DataFrame'> RangeIndex: 891 entries, 0 to 890 Data columns (total 15 columns): Column Non-Null Count -------------0 int64 survived 891 non-null 1 891 non-null int64 pclass 891 non-null object sex 3 714 non-null float64 age int64 sibsp 891 non-null parch 891 non-null int64 fare 891 non-null float64 embarked 889 non-null object class 891 non-null object 9 who 891 non-null object 10 adult_male 891 non-null bool 11 deck 203 non-null object 12 embark_town 889 non-null object 13 alive 891 non-null object 14 alone 891 non-null bool dtypes: bool(2), float64(2), int64(4), object(7) memory usage: 92.4+ KB

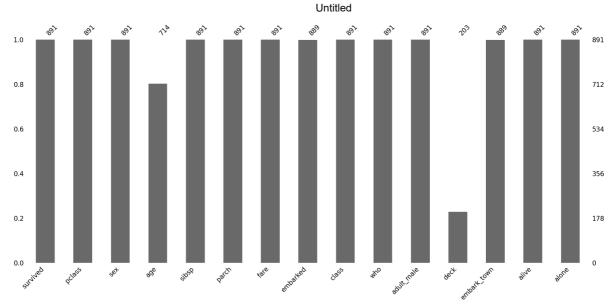
In [7]: titanic_df.isnull()

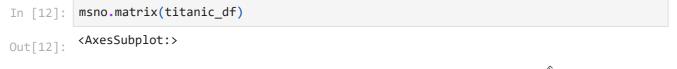
Out[7]:		survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male	deck
	0	False	False	False	False	False	False	False	False	False	False	False	True
	1	False	False	False	False	False	False	False	False	False	False	False	False
	2	False	False	False	False	False	False	False	False	False	False	False	True
	3	False	False	False	False	False	False	False	False	False	False	False	False
	4	False	False	False	False	False	False	False	False	False	False	False	True
	•••												
	886	False	False	False	False	False	False	False	False	False	False	False	True
	887	False	False	False	False	False	False	False	False	False	False	False	False
	888	False	False	False	True	False	False	False	False	False	False	False	True
	889	False	False	False	False	False	False	False	False	False	False	False	False
	890	False	False	False	False	False	False	False	False	False	False	False	True

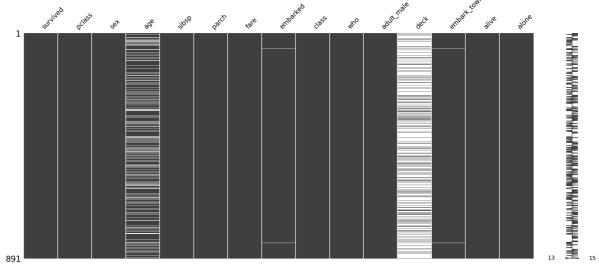
891 rows × 15 columns

2. Visualize the missing values as bar plot and matrix plot using missingno

```
In [11]: msno.bar(titanic_df)
Out[11]: <AxesSubplot:>
```







3. Handle the missing values by deleting data objects.

n [13]:	titanic_df.is	null().sum()			
ut[13]:	survived	0			
AC[TJ].	pclass	0			
	sex	0			
	age	177			
	sibsp	0			
	parch	0			
	fare	0			
	embarked	2			
	class	0			
	who	0			
	adult_male	0			
	deck	688			
	embark_town	2			
	alive	0			
	alone	0			
	dtype: int64				

```
df = titanic_df.dropna(axis=0)
In [14]:
         df.isnull().sum()
        survived
Out[14]:
                      0
        pclass
         sex
                      0
        age
                      0
                      0
        sibsp
        parch
        fare
        embarked
        class
        who
                      0
        adult_male
                      0
        deck
        embark_town
                      0
        alive
                      0
         alone
                      0
        dtype: int64
In [15]: df.info()
        <class 'pandas.core.frame.DataFrame'>
        Int64Index: 182 entries, 1 to 889
        Data columns (total 15 columns):
             Column
                        Non-Null Count Dtype
         ---
            -----
                         -----
                         182 non-null
         0
            survived
                                        int64
                        182 non-null int64
         1 pclass
         2 sex
                        182 non-null object
         3
                        182 non-null float64
            age
            sibsp
                        182 non-null
                                       int64
                       182 non-null
             parch
                                       int64
         5
            fare 182 non-null float64
embarked 182 non-null object
class 182 non-null object
         6
         7
         8 class
         9 who
                       182 non-null object
         10 adult_male 182 non-null
                                       bool
                                      object
         11 deck
                         182 non-null
         12 embark_town 182 non-null object
         13 alive 182 non-null
                                        object
         14 alone
                         182 non-null
                                        bool
         dtypes: bool(2), float64(2), int64(4), object(7)
        memory usage: 20.3+ KB
```

4. Handle the missing values by deleting attributes

```
survived
                          0
Out[17]:
                          0
         pclass
                          0
         sex
                         177
         age
         sibsp
                          0
         parch
         fare
         embarked
         class
                          0
         who
         adult_male
         embark_town
                          2
         alive
         alone
         dtype: int64
```

5. Handle the missing value by imputing the missing values with arbitrary value

```
In [18]: titanic_df['deck'].unique()
Out[18]: array([nan, 'C', 'E', 'G', 'D', 'A', 'B', 'F'], dtype=object)
In [19]: titanic_df['deck'].isnull().sum()
Out[19]: 688
In [20]: titanic_df['deck'] = titanic_df['deck'].fillna('C')
In [21]: titanic_df['deck'].isnull().sum()
Out[21]: 0
In [22]: titanic_df
```

Out[22]:		survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
	0	0	3	male	22.0	1	0	7.2500	S	Third	man	True
	1	1	1	female	38.0	1	0	71.2833	С	First	woman	False
	2	1	3	female	26.0	0	0	7.9250	S	Third	woman	False
	3	1	1	female	35.0	1	0	53.1000	S	First	woman	False
	4	0	3	male	35.0	0	0	8.0500	S	Third	man	True
	•••											
	886	0	2	male	27.0	0	0	13.0000	S	Second	man	True
	887	1	1	female	19.0	0	0	30.0000	S	First	woman	False
	888	0	3	female	NaN	1	2	23.4500	S	Third	woman	False
	889	1	1	male	26.0	0	0	30.0000	С	First	man	True
	890	0	3	male	32.0	0	0	7.7500	Q	Third	man	True
	891 ro	ws × 15	column	S								

Mean

```
In [23]:
         mean = titanic_df['age'].mean()
          print(mean)
          #Replace the missing values for numerical columns with mean
          titanic_df['age'] = titanic_df['age'].fillna(mean)
          titanic_df['age']
         29.69911764705882
                22.000000
Out[23]:
                 38.000000
         2
                26.000000
         3
                 35.000000
                 35.000000
         886
                27.000000
         887
                19.000000
         888
                29.699118
         889
                 26.000000
         890
                32.000000
         Name: age, Length: 891, dtype: float64
```

Mode

```
In [27]: #Replace the missing values for categorical columns with mode
    mode = titanic_df['deck'].mode()[0]
    print(mode)
    titanic_df['deck'] = titanic_df['deck'].fillna(mode)
    C
In [28]: titanic_df['deck']
```

```
Out[28]: 0 C
1 C
2 C
3 C
4 C
...
886 C
887 B
888 C
889 C
890 C
Name: deck, Length: 891, dtype: object
```

Median

```
In [29]: titanic_df['age']= titanic_df['age'].fillna(titanic_df['age'].median())
          titanic_df['age']
                 22.0
Out[29]:
                 38.0
                26.0
         3
                 35.0
                 35.0
         886
                27.0
         887
                19.0
         888
                28.0
         889
                26.0
         890
         Name: age, Length: 891, dtype: float64
```

6. Handle the missing value by imputing the missing values using forward fill and backward fill

```
In [31]: titanic_df = pd.read_csv("titanic2.csv")
    titanic_df
```

6/22, 1:59 PM								Untitled				
Out[31]:		survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
	0	0.0	3	male	22.0	1	0	7.2500	S	Third	man	True
	1	1.0	1	female	38.0	1	0	71.2833	С	First	woman	False
	2	NaN	3	female	NaN	0	0	7.9250	S	Third	woman	False
	3	NaN	1	female	NaN	1	0	53.1000	S	First	woman	False
	4	0.0	3	male	35.0	0	0	8.0500	S	Third	man	True
	•••											
	886	0.0	2	male	27.0	0	0	13.0000	S	Second	man	True
	887	1.0	1	female	19.0	0	0	30.0000	S	First	woman	False
	888	0.0	3	female	NaN	1	2	23.4500	S	Third	woman	False
	889	1.0	1	male	26.0	0	0	30.0000	С	First	man	True
	890	0.0	3	male	32.0	0	0	7.7500	Q	Third	man	True
4	_											
In [32]:	new_	_df = tit _df	anic_d	f.filln	a(met	hod="f	fill")					•
In [32]: Out[32]:	_	_		f.filln sex			fill") parch		embarked	class	who	adult_male
	_	_df		sex	•				embarked S	class Third	who	
	new_	df survived	pclass	sex male	age	sibsp	parch 0	fare		Third		adult_male
	new_	survived 0.0	pclass 3	sex male	age 22.0 38.0	sibsp	parch 0	fare 7.2500	S	Third First	man	adult_male True
	0 1	survived 0.0 1.0	pclass 3 1 3	sex male female	age 22.0 38.0 38.0	sibsp 1	parch 0 0 0	fare 7.2500 71.2833	S C	Third First Third	man	adult_male True False
	0 1 2	0.0 1.0	pclass 3 1 3	sex male female female	age 22.0 38.0 38.0 38.0	sibsp 1 1 0	parch 0 0 0	7.2500 71.2833 7.9250	S C S	Third First Third	man woman woman	adult_male True False False
	0 1 2	0.0 1.0 1.0	pclass 3 1 3 1	sex male female female female	age 22.0 38.0 38.0 38.0	1 1 0 1	parch 0 0 0 0	7.2500 71.2833 7.9250 53.1000	S C S	Third First Third First	man woman woman woman	adult_male True False False False
	0 1 2 3 4	0.0 1.0 1.0 1.0	pclass 3 1 3 1 3	sex male female female female male	age 22.0 38.0 38.0 38.0	1 1 0 1 0	parch 0 0 0 0 0	7.2500 71.2833 7.9250 53.1000 8.0500	S C S S	Third First Third First Third	man woman woman woman man	adult_male True False False False True
	0 1 2 3 4	0.0 1.0 1.0 1.0 0.0	pclass 3 1 3 1 3 2	sex male female female male	age 22.0 38.0 38.0 38.0 35.0 27.0	1 1 0 1 0	parch 0 0 0 0 0	7.2500 71.2833 7.9250 53.1000 8.0500	S C S S	Third First Third First Third Second	man woman woman woman man	adult_male True False False True True
	0 1 2 3 4 886	df survived 0.0 1.0 1.0 0.0 0.0 0.0	pclass 3 1 3 1 3 2 1	sex male female female male male	age 22.0 38.0 38.0 35.0 27.0	\$\frac{1}{1} \\ 0 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \	parch 0 0 0 0 0 0	fare 7.2500 71.2833 7.9250 53.1000 8.0500 13.0000	S C S S S	Third First Third First Third Second First	man woman woman man man	adult_male True False False True True
	0 1 2 3 4 886 887	df survived 0.0 1.0 1.0 0.0 0.0 1.0	pclass 3 1 3 1 3 2 1	sex male female female male male female female	age 22.0 38.0 38.0 35.0 27.0 19.0	\$\frac{1}{1} \\ 0 \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \\ \	parch 0 0 0 0 0 2	7.2500 71.2833 7.9250 53.1000 8.0500 13.0000 30.0000	S C S S S S S S S S S	Third First Third First Third Second First	man woman woman man man woman	adult_male True False False True True False
	0 1 2 3 4 886 887 888	df survived 0.0 1.0 1.0 0.0 0.0 1.0 0.0 0.0	pclass 3 1 3 1 3 2 1 3	sex male female female male male female female	age 22.0 38.0 38.0 35.0 27.0 19.0 26.0	\$\frac{1}{1} \\ 0 \\ \dots \dots \\ \do	parch 0 0 0 0 0 2	fare 7.2500 71.2833 7.9250 53.1000 8.0500 13.0000 30.0000 23.4500	S C S S S S S S	Third First Third First Third Second First Third	man woman woman man man woman woman	adult_male True False False True True False False False
	0 1 2 3 4 886 887 888 889 890	df survived 0.0 1.0 1.0 0.0 0.0 1.0 0.0 1.0 1.0 1.0	pclass 3 1 3 1 3 2 1 3 1 3 1 3	sex male female female female male male male female female	age 22.0 38.0 38.0 35.0 27.0 19.0 26.0	\$\frac{1}{1} \\ 0 \\ \dots \dots \\ \do	parch 0 0 0 0 0 0 2 0	fare 7.2500 71.2833 7.9250 53.1000 8.0500 13.0000 30.0000 23.4500 30.0000	S C S S S C C	Third First Third Second First Third First	man woman woman man man woman woman man	adult_male True False False True True False False True True True False True

```
In [33]: new_df = titanic_df.fillna(method="bfill")
new_df
```

Out[33]:		survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_male
	0	0.0	3	male	22.0	1	0	7.2500	S	Third	man	True
	1	1.0	1	female	38.0	1	0	71.2833	С	First	woman	False
	2	0.0	3	female	35.0	0	0	7.9250	S	Third	woman	False
	3	0.0	1	female	35.0	1	0	53.1000	S	First	woman	False
	4	0.0	3	male	35.0	0	0	8.0500	S	Third	man	True
	•••											
	886	0.0	2	male	27.0	0	0	13.0000	S	Second	man	True
	887	1.0	1	female	19.0	0	0	30.0000	S	First	woman	False
	888	0.0	3	female	26.0	1	2	23.4500	S	Third	woman	False
	889	1.0	1	male	26.0	0	0	30.0000	С	First	man	True
	890	0.0	3	male	32.0	0	0	7.7500	Q	Third	man	True

891 rows × 15 columns

In []: